STUDY ON SPATIO-TEMPORAL PROPERTIES

OF RAINFALL

A Dissertation

by

JANGHWOAN CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Civil Engineering

STUDY ON SPATIO-TEMPORAL PROPERTIES

OF RAINFALL

A Dissertation

by

JANGHWOAN CHOI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Francisco Olivera
Committee Members,	Anthony Cahill
	Scott Socolofsky
	Raghavan Srinivasan
	Courtney Schumacher
Head of Department	David Rosowsky

December 2006

Major Subject: Civil Engineering

ABSTRACT

Study on Spatio-Temporal Properties of Rainfall. (December 2006)

Janghwoan Choi, B.S.; M.S., Inha University, Incheon, South Korea

Chair of Advisory Committee: Dr. Francisco Olivera

This dissertation describes spatio-temporal properties of rainfall. Rainfall in space was modeled by a precipitation areal reduction factor (ARF) using a NEXRAD image. The storms are represented as ellipses, which are determined by maximizing the volume of rainfall. The study investigated 18,531 storms of different durations that took place in different seasons and regions of Texas. Statistical analysis was carried out to find a relationship between ARFs and predictor variables (storm duration, area, season, region, and precipitation depth).

The stochastic model for temporal disaggregation of rainfall data was evaluated across Texas. The hourly historic data from the selected 531 hourly gauges in Texas were used to evaluate the model's performance to reproduce hourly rainfall statistics. Spatial trends in performance statistics or spatial patterns among gauge characteristics (e.g. period of record, precipitation statistics) were examined by cluster analysis. Since no spatial trends or patterns were identified, the state database is used and verified for a selection of gauges. The method was further applied to estimate intensity-duration curves for hydrologic applications.

To obtain basic information on the spatial and dynamic patterns of rainfall over an area, it is necessary to identify and track a storm objectively. Automated algorithms are needed to process a large amount of radar images. A methodology was presented to overcome the identification and tracking difficulties of one-hour accumulated distributed rainfall data and to extract the characteristics of moving storms (e.g., size, intensity, orientation, propagation speed and direction, etc.). The method presented in this dissertation allows the user to better understand the precipitation patterns in any given area of the United States, and yields parameters that describe storm dynamic characteristics. These parameters can then be used in the definition of synthetic dynamic storms for hydrologic modeling.

DEDICATION

This dissertation is dedicated to God, who walks with me every step of the way.

This dissertation is also dedicated to my family.

ACKNOWLEDGMENTS

First and foremost, I would like to thank God for granting me the strength, determination, courage, and wisdom to do all things through Jesus Christ.

I would like to gratefully and sincerely thank my advisor, Francisco Olivera, for his guidance, support and patience during my Ph.D. program. Many thanks also to committee members Tony Cahill, Scott Socolofsky, Raghavan Srinivasan, and Courtney Schumacher.

I would like to express my sincere appreciation to the people who have assisted in so many ways before and during my work at Texas A&M. It is not possible to enumerate everybody who supported me; however in particular, I would like to thank Huidae Cho and Dongkyun Kim at Texas A&M University.

I send my thanks to my family for their constant support and strong belief in me. I would like to express my deepest gratitude to my wife, Sueun Kim, for all her love, companionship and encouragement. You were there for me whenever I needed you. I love you from the bottom of my heart.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

Page

1. INTRODUCTION

Flow prediction and forecasting requires the understanding of the spatio-temporal characteristics of rainfall fields. The availability of a relatively dense rain gauge network combined with precipitation images derived from radar reflectivity nationwide has created excellent opportunities to investigate rainfall patterns in space and time.

This dissertation presents methods for quantifying spatial and temporal characteristics of rainfall fields. Three applications are described in detail. First, areal precipitation estimates were calculated with *Areal Reduction Factors* (ARFs), which are numbers that multiply point rainfall to give expected average rainfall over a given area. Second, disaggregation of observed daily precipitation into hourly precipitation over Texas was conducted using a stochastic disaggregation technique (Socolofsky et al. 2001). Finally, a storm feature identification and tracking algorithm is presented and used to analyze their dynamic patterns over Brazos County in Texas. The first and third applications use the one-hour-accumulated 4-km-by-4-km averaged next generation radar (NEXRAD) rainfall data (NWS 2005a) to describe precipitation over time and space. The second application uses the rain gauge network of the National Climate Data Center (NCDC) (NCDC 2003) for Texas.

1.1. SPATIO-TEMPORAL VARIABILITY OF RAINFALL

Estimated watershed responses to rainfall events are strongly influenced by the spatio-temporal variability of precipitation (Alley 1981; Alley and Smith 1981; Donigian et al. 1984; Krajewski et al. 1993; Ogden and Julien 1993; Ogden et al. 1995; Goodrich et al. 1995; Marani et al. 1997; Singh 1997, 1998; Chaubey et al. 1999; Zeng et al. 2000; Margulis and Entekhabi 2001; Güntner et al. 2001; de Lima and Sigh 2001).

This dissertation follows the style of *Journal of Hydrologic Engineering.*

Wilson et al. (1979) showed that small-catchment hydrographs are affected by the spatial distribution of rainfall. Likewise, Krajewski et al. (1993) indicate that hydrologic models that use spatially-variable rainfall give better results than those that use uniform rainfall. Chaubey et al. (1999), additionally, studied the effect of the spatial variability of rainfall on the uncertainty of the estimated model parameters, and showed that the assumption of spatially-uniform rainfall can lead to large uncertainties.

Accurate representation of the temporal variability of rainfall (i.e., intensity and storm intermittency) is critical to correctly predict runoff (Alley 1981; Alley and Smith 1981; Donigian et al. 1984; Marani et al. 1997; Zeng et al. 2000; Margulis and Entekhabi 2001; Güntner et al. 2001). Zeng et al. (2000), for example, found that interception losses are controlled by the temporal variability of rainfall (i.e., the mean inter-storm time and the mean storm duration). Güntner et al. (2001) showed that, even for short storm durations, the temporal variability of rainfall has a significant effect on runoff hydrographs and flood frequency curves.

Ogden and Julien (1993) examined the sensitivity of runoff depths to the rainfall spatial variability for different storm durations and sampling intervals. Their results showed that runoff sensitivity to the precipitation spatial variability decreases as the storm duration increases, and increases as the sampling interval increases. Goodrich et al. (1995) showed that rainfall variability in space and time affects precipitation estimates even over areas as small as 5 hectares and time intervals as short as 5 minutes.

The effect of storm movement on the shape and peak of runoff hydrograph has been investigated by a number of researchers (Ogden et al. 1995; Singh 1997, 1998; de Lima and Singh 2001; among others). Singh (1997, 1998), in particular, studied the influence of storm direction, areal coverage and duration on watershed responses. De Lima and Singh (2001) applied a non-linear kinematic wave model to numerically simulate flows generated by synthetic moving storms varying in storm patterns (rainfall intensity, rainfall volume and size) and storm motion (direction and speed). They found that hydrologic responses are strongly influenced by the rainfall volume, precipitation intensity, and the speed and direction of movement.

The recognition of the importance of the spatio-temporal variability of rainfall has led to the development of models that address these characteristics.

1.1.1. Average Areal Precipitation Estimates

The estimation of average areal precipitation is important for the cost-effective design of hydraulic structures that drain large areas. Areal rainfall estimates have been obtained traditionally by spatial interpolation techniques based on the assumption that rain at ungauged sites can be obtained as the weighted average of observed gauges in its vicinity (e.g., Thiessen polygons). To account for the fact that precipitation intensity averaged over large areas tends to be lower than at individual points, depth-area-duration (DAD) curves have been developed. These curves provide factor values lower than one to multiply the point precipitation and apply this resulting precipitation throughout the area. To account for the fact that precipitation intensity over long periods is not constant, hyetographs are used to represent the temporal variability of a storm.

For use as input to rainfall runoff models of large drainage areas, point rainfall depth for a given duration and return period would not be adequate to represent the average precipitation depth. High intensity rainfall at one point is unlikely to take place consistently over a large area (Siriwardena and Weinmann 1996). Hydrologic modeling of a watershed too large to sustain uniform distribution of rainfall and with a sparse network of rain gauges would require of relationships of point precipitation and area-average precipitation for areal rainfall estimation. The average areal rainfall estimates can be obtained with the ratios of areal average precipitation to point precipitation, also called Areal Reduction Factors.

Historically, estimates of areal rainfall have been obtained largely by using a network of rain gauges. While rain gauges provide the primary source of rainfall estimates, the existing network of rain gauges has several problems and limitations largely associated with poor representation of the rainfall spatial variability. A dense network of rain gauges is necessary to capture the significant spatial and temporal variability of precipitation in regions where convective storms prevail (Goodrich et al 1995; Wheater et al. 2000). Wheater et al. (2000) found that a relatively dense network of rain gauges with approximately 10 km inter-gauge spacing often failed to capture spatial heterogeneity of mesoscale convective storms, and floods occur even when no rainfall is observed. Schumacher and Johnson (2005) observed that approximately 65% of the extreme rain events in the United States during 1999-2003 were associated with mesoscale convective storms. With the advent of radar technology, spatially denser rainfall data has become available. High resolution rainfall data provide a way to characterize spatio-temporal patterns of rainfall. Taking advantage of the NEXRAD rainfall data, ARFs are developed in Texas based on the storm characteristics (depth, area, orientation, and duration), and further studied the effects of season and region on ARFs. The comprehensive literature review and methodology are presented in Section 2.

1.1.2. Rainfall Disaggregation

Continuous rainfall runoff simulations require high temporal resolution precipitation data to account for storm duration, rainfall intensity distribution over time, infiltration rates, and changes in soil moisture, among others. Despite the need of hourly rainfall data, though, there are a limited number of rain gauges that record hourly precipitation and most of them record total daily precipitation depths. In fact, historic daily rainfall gauges are nearly three times denser than hourly gauges in the United States (Bonner 1998). Thus, a model to disaggregate daily precipitation data into synthetic hourly precipitation time series has been developed. It is assumed that for the disaggregated data to be effective they must only match the rainfall statistics and measured daily totals at the disaggregation site and not the actual time of rainfall.

Two broad categories of methods are available to obtain rainfall with the correct statistics: methods based on stochastic simulation of precipitation, and deterministic methods based on direct measurement of daily precipitation depth from an off-site gauge. Stochastic simulation time series are generally obtained by continuous-simulation stochastic rainfall models or through fractal random cascade models. Most stochastic rainfall models are based on the Bartlett-Lewis or Newman-Scott rectangular pulses models. The advantage of synthetic data generated from the stochastic models is that they are of a continuous nature, able to be re-sampled at any desired aggregation level. For non-stochastic method, various fractal random cascade models capture the variability both between storms and within storms and require the fitting of scaling laws to apportion rainfall at different scales.

For the stochastic simulation methods, two types of disaggregation are identified: downscaling (Koutsoyiannis and Onof 2001) or direct disaggregation (Gyasi-Agyei 2005; Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003). Downscaling method estimates the model parameters for measured data at one time scale (e.g. daily) and used to obtain simulated rainfall at a lower (e.g. hourly) time scale (see e.g. Bo et al. 1994). Direct disaggregation method uses the synthetic data to reproduce measured daily rainfall totals by either sampling or conditioning the simulated rainfall (Glasbey et al. 1995; Gyasi-Agyei 2005; Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003). As an alternative to stochastic methods, deterministic rainfall disaggregation utilizes known weather patterns or measurements from nearby gauges.

Socolofsky et al. (2001) introduced a one-parameter stochastic selection algorithm to simulate daily totals using measured storm intensity patterns at an off-site hourly recording database and has not been applied outside of a limited study in Massachusetts. Socolofsky et al.'s

(2001) method is applied across Texas using measured historic hourly precipitation data for model validation and performance evaluation. A comprehensive literature review and methodology are presented in Section 3.

1.1.3. Storm Dynamics: Storm Identification and Tracking Algorithm

Despite the fact that moving storms are more the rule than the exception, hydrologic applications often simplify storm dynamics by assuming that rainfall is uniformly distributed over the entire watershed area for a given duration. However, determining the dynamic properties of storms can provide valuable information for stochastically developing synthetic storm events for hydrologic simulation models. Automated algorithms for identifying and tracking storms are necessary to determine the dynamic characteristics of moving storms from a large amount of spatially distributed precipitation data.

Rainfall is associated with complex physical processes that govern atmospheric motion. Thus, the spatio-temporal patterns of rainfall have often been described by stochastic models since they generally require relatively fewer model parameters and can be validated if large amounts of rainfall data at high spatial and temporal resolution are available. Precipitation radar images (e.g., NEXRAD rainfall data) do provide these types of data, which are used to construct stochastic continuous models of rainfall fields.

Rainfall field exhibits organized structures. Many individual thunderstorms (called rain cells) become organized into mesoscale convective system (MCSs) associated with very heavy precipitation as well as severe convective weather. The stochastic approach of modeling the hierarchical structure of rainfall fields was introduced by Le Cam (1961), who also proposed the utility of using cluster-point processes to explain the spatial and temporal characteristics of storms. Cluster point processes require a large set of parameters to represent the hierarchical structure of rainfall fields in different scales. Most spatio-temporal rainfall models require detailed information of the distributed characteristics of storm structures to generate synthetic storms.

For distributed rainfall data to be used to develop stochastic models, a means is needed to classify the rainfall structures unambiguously. In order to obtain basic information on the spatial and dynamic patterns of rainfall over an area, it is necessary to identify and track a storm objectively, and automated algorithms are needed to process a large amount of radar images. To date, most algorithms for identifying and tracking storm characteristics have been applied to short time-step radar reflectivity data (e.g., 15 minutes or less), where storm features are captured in an effectively synoptic manner. Throughout the United States, however, the most reliable data for distributed precipitation are the one-hour accumulated NEXRAD data of the U.S. National Weather Service (Klazura and Imy 1993; Brown and Lewis 2005; National Weather Service 2006; among others). The one-hour aggregation level of the data makes it more difficult to identify and track storms than when using sequences of synoptic radar reflectivity data because storms can traverse over a number of NEXRAD pixels and can change size and shape appreciably between consecutive data maps. In this dissertation, a methodology is presented to overcome these identification and tracking difficulties and to extract the characteristics of moving storms from one-hour accumulated distributed rainfall data. A comprehensive literature review and methodology are presented in Section 4.

1.2. DISSERTATION ORGANIZATION

This dissertation consists of five sections: an introductory section, which is this section; three standalone sections that cover three somewhat independent topics; and one section that discusses topics for further research. The second, third and fourth sections include their own introduction, literature review, methodology, application and conclusion sections, but not a references section, which is common for the entire dissertation. These three sections have been formatted as journal articles. Additionally, an appendix that describes in detail the methods used is presented at the end.

Section 2 presents precipitation areal reduction factors (ARFs) derived from NEXRAD rainfall estimates in Texas. A new technique is used to take into account anisotropic properties of storm events by identifying the areas producing maximum volume of rainfall, and by representing it as an elliptical shape. Several important factors are investigated to explain ARF variability by using statistical analysis. The empirical ARFs obtained from this study are compared to previous studies.

Section 3 presents a method to disaggregate daily rainfall into hourly precipitation, and evaluates the presented method using data from hourly rain gauges across Texas. Performance of the model is evaluated by the model's ability to regenerate hourly rainfall statistics. Cluster analysis is used to find spatial patterns of model performance and gauge characteristics. The method is further applied to estimate intensity duration curves for hydrologic applications.

Section 4 presents new storm feature identification and tracking algorithm using one-hour accumulated NEXRAD precipitation images. The algorithm is applied to find storm feature characteristics (i.e., size, shape, direction and speed of moving storm) and the spatial and dynamic patterns of the storm features over Brazos County in Texas of year 2003. The statistical results of the storm feature characteristics are summarized and discussed.

2. ESTIMATION OF AVERAGE RAINFALL AREAL REDUCTION FACTORS IN TEXAS USING NEXRAD DATA

Precipitation areal reduction factors (ARFs) for the 685,000 km^2 of Texas were calculated using NEXRAD rainfall estimates. The study was based on 18,531 storms of different durations that took place in different seasons and regions of Texas. The rainfall field was considered anisotropic, and the storms were assumed of elliptical shape. It was found that, in addition to the storm duration and area, other factors such as the season, region and precipitation depth have a statistically significant effect on the ARFs. Elongated ellipses and orientation angles somewhat parallel to the Texas gulf coast were found more frequent in winter, when warm and cold fronts produce frontal storms, than in summer. The effect of the precipitation depth on the ARFs was found to be stronger in summer than in winter. Even though part of the ARF variability could be explained by seasonality, regionality and precipitation depth, the uniqueness of each storm event appears to be an important cause of it. Lower ARF values were observed compared to previous studies.

2.1. INTRODUCTION

In general, "larger catchments are less likely than smaller catchments to experience high intensity storms over the whole of the catchment area" (Siriwardena and Weinmann 1996). Therefore, the conversion of point precipitation into area-averaged precipitation is necessary whenever an area, large enough for rainfall not to be uniform, is being modeled. However, while point precipitation has been well recorded with rain gauges, areal precipitation cannot be measured and its estimation has been subject of research for the last decades (Weather Bureau 1957, 1958a, 1958b, 1959, 1960, 1964; Rodriguez-Iturbe and Mejia 1974; Frederick et al. 1977; NWS 1980; Omolayo 1993; Srikathan 1995; Bacchi and Ranzi 1996; Siriwardena and Weinmann

1996; Sivapalan and Bloschl 1998; Asquith and Famiglietti 2000; De Michele et al. 2001; Durrans et al. 2002; among others).

With the understanding that the NEXt generation RADar (NEXRAD) (NWS 2005a) precipitation estimates are the best spatially distributed rainfall data for large areas in the United States, this dissertation addresses the calculation of the ratios of areal precipitation to point precipitation – also called areal reduction factors (ARFs) – in the 685,000 km^2 of Texas using these data. Because of the continuous improvements of the NEXRAD data, which cause them not to be homogeneous over time, only data for years 2003 and 2004 were considered in the study. This lack of long term historical data prevented any type of frequency analysis and, instead, ARFs were related to rainfall depths for given durations (which are associated to specific return periods). Still, the use of only two years of data offered a relatively narrow range of precipitation depths in which low frequency values were not included. The analysis was based on 18,531 snapshots of storms that took place in the study area in the two year period.

It was found that, in addition to the storm duration and area, already considered in previous studies, ARFs also depend on the season and geographic region in which the storm takes place, and on the precipitation depth. Additionally, a discussion on the most likely storm shape (i.e., aspect ratio of an assumed elliptical storm area) and orientation angle for different seasons and regions is included.

2.1.1. Background

A number of approaches for converting point precipitation into areal precipitation are based on precipitation records. Traditionally, ARF estimation algorithms have been grouped under two broad categories: those based on rain gauge networks (known as geographically fixed), and those based on individual storm events (known as storm centered) (Srikathan 1995). The geographically fixed approach is particularly suited for discrete (i.e. point) precipitation data and

ARFs are estimated using data of rain gauge networks. ARFs are calculated as the ratio of a representative precipitation depth over the area covered by the network to a representative point precipitation depth. How to estimate the area covered by the network, how to estimate the representative precipitation depth over the area, or how to estimate the representative point precipitation depth, however, change from method to method. Moreover, based on the algorithms used to calculate geographically fixed ARFs, it can be said that they do not consider concurrent precipitation depths (i.e., the areal and point precipitation do not correspond necessarily to the same event), and they are sensitive to the configuration of the network (i.e., adding or removing a rain gauge affects the ARF values). The storm centered approach, on the other hand, is suited for continuous (i.e., surface) precipitation data, such as radar data. In this case, ARFs are calculated for individual events for which they describe their areal properties, and are equal to the ratio of the average precipitation depth over a given area to the concurrent point precipitation depth in the storm center. Because storm centered ARFs are estimated for individual events, they can capture the anisotropy of the rainfall field. The storm centered approach, however, has the disadvantage that the ARFs are "applicable to specific types of storm events" (Srikathan 1995); and, therefore, unless a large sample of storm centered ARFs are estimated, representative values would not be captured. Broadly speaking, geographically fixed ARFs are the result of averaging precipitation data and then calculating ARF values, while storm centered ARFs are the result of calculating ARF values for each of a large sample of storms and then averaging them. Both approaches have strengths and weaknesses, and, in both cases, the application of their results to areas different from those for which they were derived should be done carefully.

One of the first attempts at estimating ARFs in the United States was conducted by the Weather Bureau (1957, 1958a, 1958b, 1959, 1960) in Technical Paper 29, frequently referred to as TP-29. In TP-29, for the area associated with a precipitation gauge network, the ARF values were estimated as the ratio of the average annual maximum areal precipitation depth for the analysis period to the average annual maximum point precipitation depth in the area for the analysis period. Because of the averaging of the areal and point precipitation depths, the resulting factors correspond to a return period of 2.33 years. Moreover, according to TP-29, ARFs do not depend on geographic location, which implies that the same values apply regardless of the local climate conditions. Later, in Technical Paper 49 (i.e., TP-49) (Weather Bureau 1964), the relationship between the ARFs and the storm frequency was studied, and it was concluded that ARFs are not affected by the recurrence interval. Additionally, because of the insufficient precipitation data available, no clear indication was found to relate ARFs to geographic location. Omolayo (1993), in turn, studied the applicability of the ARF values of TP-29 to watersheds in Australia, and concluded that they are "probably satisfactory … for estimating 24-hour area[averaged] rainfalls for area sizes between 200 and 500 km^2." Omolayo (1993) also indicates the inadequacy of using point rainfall for estimating areal rainfall of a particular frequency with the storm centered approach. Likewise, as part of an effort for calculating ARFs for the region of Victoria in Australia, Siriwardena and Weinmann (1996) estimated ARF values for a large number of circular sample catchments based on daily rainfall data. Although they found "small, but statistically significant, differences in ARF values for different parts of Victoria", there was insufficient information to recognize patterns in the ARFs based on geographic location. Siriwardena and Weinmann (1996) also found that ARFs tend to increase with the recurrence interval. Similarly, Asquith and Famiglietti (2000) proposed an "annual maxima centered approach" for estimating ARFs assuming an isotropic precipitation field. ARF values for one-day storms for Austin, Dallas and Houston in Texas were estimated, and the resulting values were found to be lower than those proposed in TP-29. For their dataset, Asquith and Famiglietti (2000) found dependency of the ARFs on geographic location and increasing values as the storm recurrence interval increased.

Analytical approaches for converting point precipitation into areal precipitation have also been developed. A frequent assumption in these analytical approaches is the representation of the precipitation field as isotropic; that is, the spatial structure of the precipitation field depends only on distance and not on direction. Rodriguez-Iturbe and Mejia (1977) proposed a methodology in which the ARFs depend solely on the expected correlation coefficient between the precipitation depths at two randomly chosen points. Sivapalan and Blöschl (1998), likewise, proposed a method for constructing intensity-duration-frequency (IDF) curves consisting of the areal averaging of the parent rainfall and the transformation of the parent rainfall distribution into an extreme value distribution. Similarly, after obtaining a scaling relation of average rainfall intensity in area and duration, De Michele et al. (2001) calculate ARFs of extreme rainfall events based on dynamic scaling and statistical self-affinity.

With the advent of radar technology, spatially denser precipitation data has become available. Frederick et al. (1977) states that, since ARFs estimated from radar-based precipitation data are expressed as "dimensionless ratios of areal average precipitation to point precipitation, both being estimated from radar digits, […] the uncertainties [in the] Z-R relationships are in part 'divided out'." Likewise, Bacchi and Ranzi (1996) present a stochastic derivation of ARFs "based on the analysis of the crossing properties of the rainfall process aggregated in space and time." They state that "radar data are more efficient than usual rain gauge networks in capturing the internal structure and the spatial distribution of storms" and that for estimating ARFs "not actual intensity values, but the ratio between areal and point intensities, are needed." Bacchi and Ranzi (1996) also indicate that "the use of radar data should give at least as reliable results as those achievable by using only gauge data, unless very dense and large [precipitation gauge] networks are available." Durrans et al. (2002), in turn, evaluated ARFs for the Arkansas-Red River basin using the methodology of TP-29 and NEXRAD data. Their results were "consistent" with TP-29, although differences clearly exist. According to Durrans et al. (2002), "the most significant limitations of radar rainfall data, both for frequency analysis and for development of depth-area relationships, are the shortness of the records and the heterogeneities caused by continual improvements to the data processing algorithms."

In this dissertation, the estimation of ARFs in Texas using NEXRAD data is presented. ARFs, storm shapes and orientations were determined for different storm durations, areas, seasons, regions and precipitation depths. Because the anisotropy of the storm field played an important role in the estimation of ARFs, the storm-centered ARF calculation method was used. It is apparent that the estimation of ARFs is a complex problem, and that more research in the light of more and better data will always be necessary. This study aims at adding a new perspective to the already rich discussion on ARF estimation.

2.1.2. NEXRAD Data

In the United States, nationwide radar-based precipitation depths are estimated as part of the NEXRAD federal program and distributed by the National Weather Service (NWS). This program has resulted in the delivery of a number of S-band Weather Surveillance Radar-1988 Doppler (WSR-88D) radars across the country (Fulton et al. 1998). There are 12 WSR-88D radars in Texas plus several close to the state border, which cover most of the study area with the exception of part of the Big Bend country (Figure 2-1).

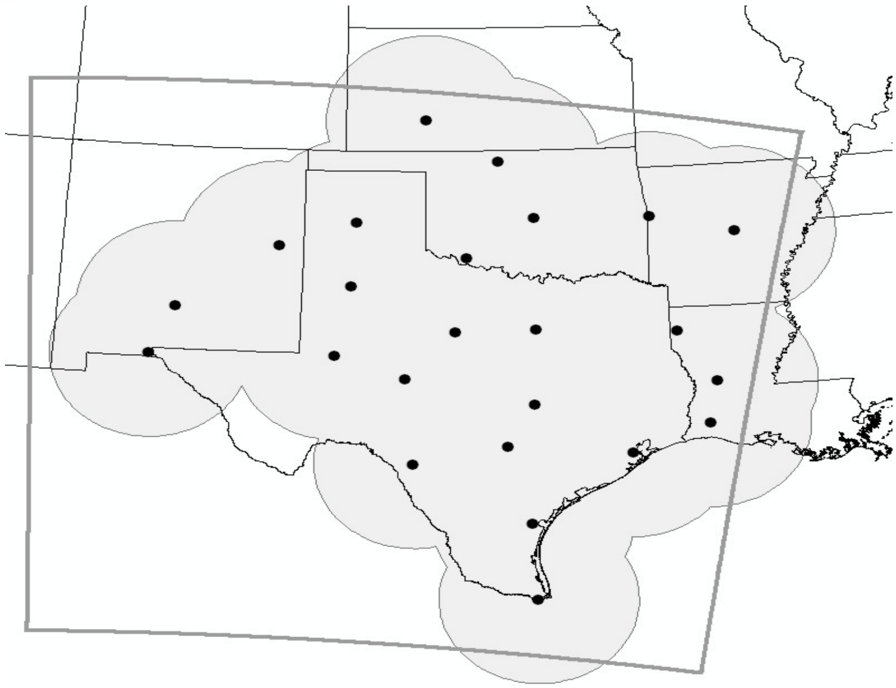


Figure 2-1. Radar locations (black points) and scanning domain (gray area) within and around Texas, and West Gulf River Forecasting Center (WGRFC) HRAP grid (thick black line).

NEXRAD rain estimates are distributed in hourly digital precipitation arrays (DPAs) mapped onto the national Hydrologic Rainfall Analysis Project (HRAP) grid. The HRAP grid cell size is nominally 4 km, but ranges from about 3.7 km at southern U.S. latitudes to about 4.4 km at northern U.S. latitudes (Fulton 1998). In Texas, it ranges from 3.7 km to 4.1 km. Because of the different sources of data used to develop the NEXRAD precipitation dataset (i.e., radar, gauges and satellite imagery), it is called Multisensor Precipitation Estimator (MPE). Detailed discussion on the development of radar-based precipitation depth maps, sources of error and algorithms used for adjusting precipitation estimates to observed data is presented in Fulton et al. (1998), Young et al. (2000), NWS (2002), McCollum et al. (2002), Krajewski and Smith (2002), Houze et al. (2004) and Gebremichael and Krajewski (2004), among others.

Not withstanding the fact that NEXRAD precipitation data are subject to inaccuracies due to a number of error sources (NWS 2002), it should be acknowledged that, at present, it is the best continuous spatially-distributed precipitation database available statewide (and country-wide) at its resolution.

2.2. METHODOLOGY

The hourly DPAs of years 2003 and 2004 of the West Gulf River Forecasting Center (WGRFC) HRAP grid (Figure 2-1) were obtained from the NWS Internet site (NWS 2005b).

2.2.1. Selection of Storm Cells

The WGRFC HRAP grid has 165,750 cells (390 rows × 425 columns), and covers an area significantly larger than the state of Texas (Figure 2-1). Therefore, to limit the calculations to the study area, the grid was clipped out by a polygon that buffered Texas around by 50 km. This clipping decreased the number of cells to 56,420.

For each cell of the clipped grid, the annual maximum precipitation depth was calculated for storm durations of 1, 3, 6, 12 and 24 hours, and its value was stored along with its corresponding time of occurrence. For durations other than one hour (which is the time resolution of the original precipitation dataset), a moving window over time was defined so that the annual maximum accounted for the consecutive hours that generated the maximum depth for the given duration. As a result, grids of annual maxima and dates of occurrence were generated for the different durations. Even though the methodology presented here could have been applied to any storm event, the calculations were limited to annual maxima because high-precipitation storms are less frequent and were necessary to describe a wider range of precipitation depths.

Likewise, only cells with precipitation depths greater than 20 mm, 25 mm, 30 mm, 35 mm and 40 mm for durations of 1, 3, 6, 12 and 24 hours, respectively, were considered. These threshold depths correspond approximately to the minimum value of the two-year precipitation depth in Texas (Weather Bureau 1961).

Additionally, ARFs were calculated only for those cells in which the precipitation value was the greatest in the 21×21-cell window of concurrent precipitation centered at the cell (Figure 2-2). The 21×21-cell window size was set to fit the largest storms analyzed, which were of around 800 km^2 and (assuming elliptically-shaped storms) five times longer than wider.

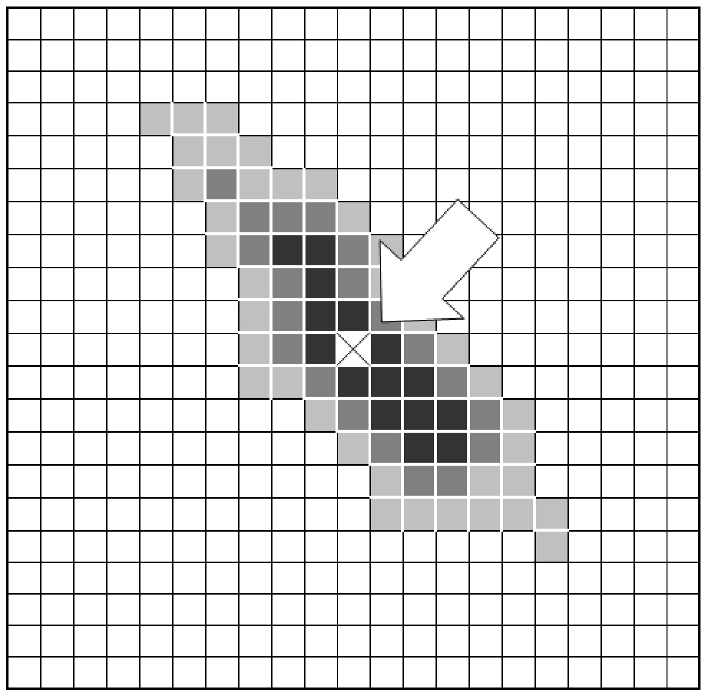


Figure 2-2. 21×21-cell window centered at the cell being studied. ARFs are calculated only if the center cell has the maximum precipitation depth in the window.

Thus, ARFs were calculated for 7,479 storm snapshots of 1-hour duration, 4,189 of 3-hour duration, 2,895 of 6-hour duration, 2,173 of 12-hour duration and 1,795 of 24-hour duration for a total of 18,531. Figure 2-3 shows the location of the storms for the specific case of a duration of 1 hour. Overall, the storms were well distributed over the study area with the exception of part of the Big Bend country, which is related to the lack of radar coverage in the area.

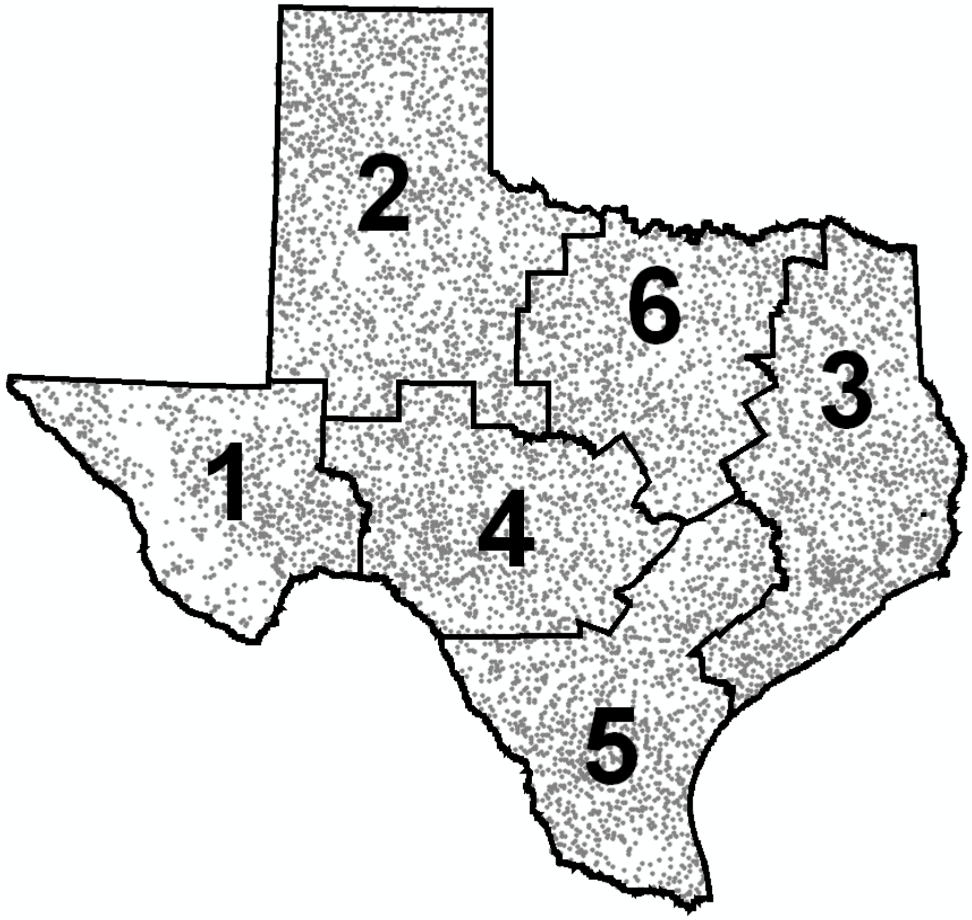


Figure 2-3. Location of the cells for which ARFs were calculated for a storm duration of 1 hour.

2.2.2. ARF Calculation

The storms were assumed to be of elliptical shape and centered at the selected cells. To determine the storm ellipse for a given area, its aspect (i.e., the ratio of its long to its short diameter) and orientation (i.e., the angle formed by its long diameter and the east axis) were changed systematically, until the volume inside the ellipse was maximized (Figure 2-4). In the following, these ellipses are called optimum ellipses.

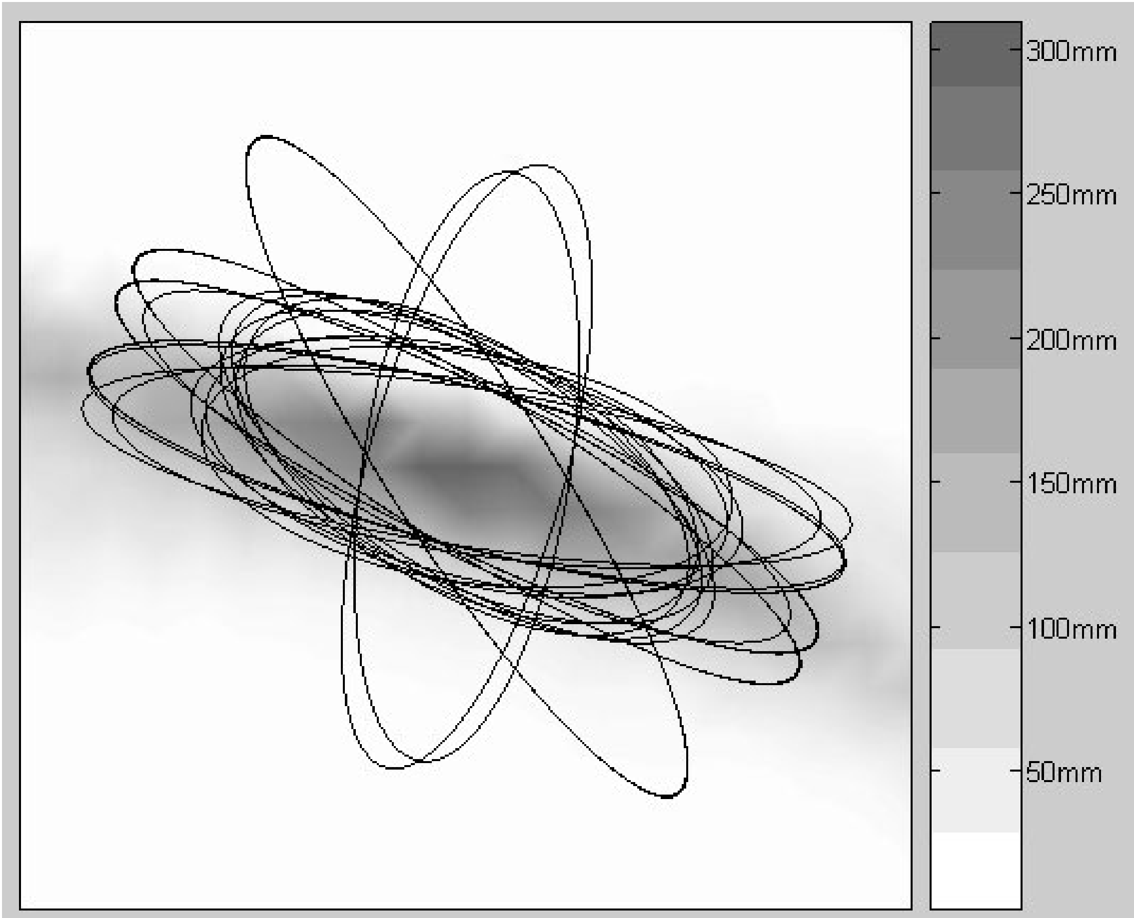


Figure 2-4. The optimum ellipse for a given area was found by systematically changing the ellipse aspect ratio and orientation. The scale on the right corresponds to the precipitation depth.

ARFs were then calculated according to

$$ARF_{A_o} = \frac{\frac{1}{A}\iint\limits_{A_o} P\,dA}{P_o} \qquad (2\text{-}1)$$

where ARF_{A_o} is the areal reduction factor for a given cell, optimum ellipse area and storm duration; $A_o[L2]$ is the area of the optimum ellipse; P $[L]$ is the precipitation depth at the points inside the optimum ellipse; and P_0 $[L]$ is the precipitation depth at the center cell. Note that there are several optimum ellipses of different areas centered at the same cell (Figure 2-5); thus, producing different ARFs for different areas.

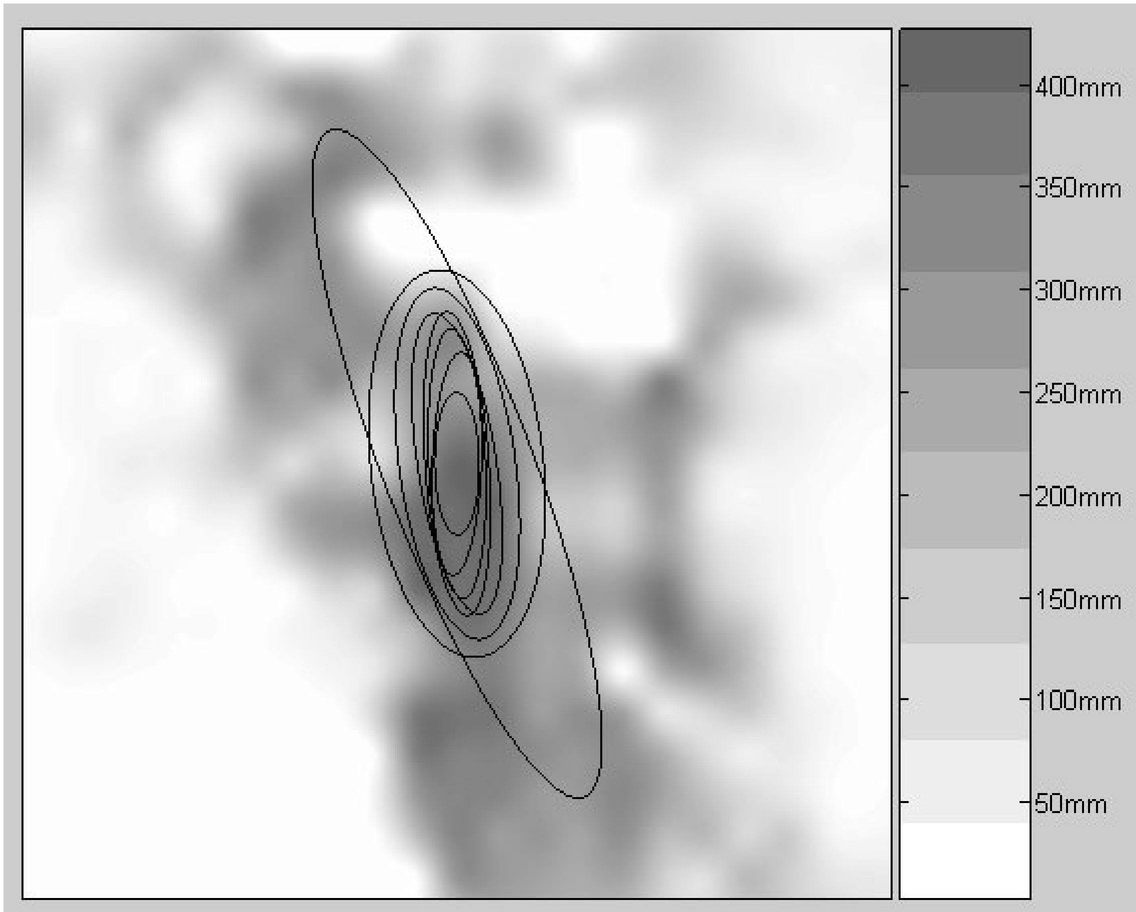


Figure 2-5. Optimum ellipses of different areas for a given cell. Each ellipse has its own aspect ratio and orientation angle. The scale on the right corresponds to the precipitation depth.

Equation (2-1) was used to calculate ARFs of 148,248 optimum ellipses (i.e., 18,531 cells × 8 ellipses/cells) of average areas of 45, 75, 105, 135, 166, 286, 467 and 737 km^2. The result of the ARF calculations was a 148,248-row and a 9-column table, in which each row corresponded to an optimum ellipse, and the columns were the cell's X and Y coordinates, the storm's duration, precipitation depth in the cell and time of occurrence, the ellipse's area, aspect ratio and orientation angle, and the ARF value.

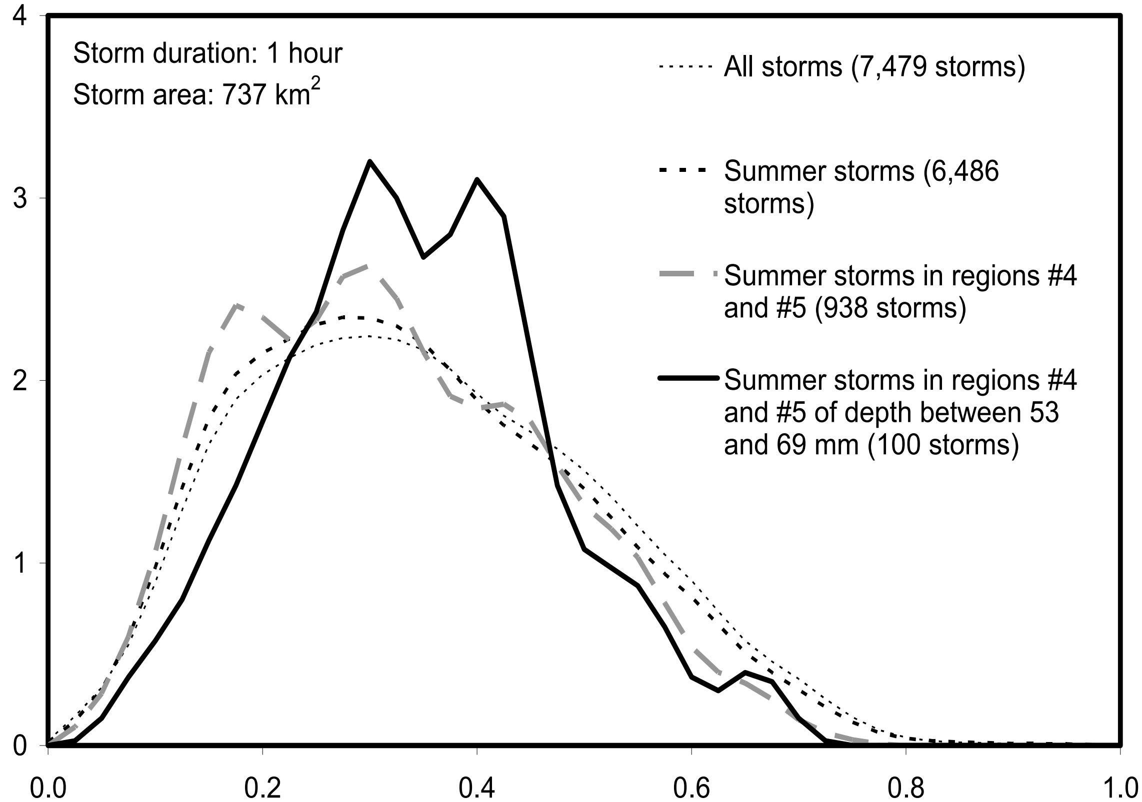


Figure 2-6. ARF variability for storms of duration of 1 hour and area of 737 km^2 in Texas.

Previous studies show that ARFs decrease with area and increase with storm duration; however, their variability with time of occurrence (seasonality), location (regionality) and precipitation depth (associated with the storm frequency) has not been investigated with the same level of detail. The “All storms” curve in Figure 2-6 shows the frequency histogram of ARFs for a duration of 1 hour and area of 737 km^2. The variability of the values suggests that factors, other than duration and area, also affect the ARFs.

2.2.3. Effect of the Storm Season and Region on the ARFs

Expecting the storms to be driven by different atmospheric processes in different parts of the year, the year was subdivided into two six-month seasons: summer and winter. Broadly speaking, in summer storms, the air lifting necessary for air moisture condensation is driven by the vertical temperature gradient (i.e., convective lifting); whereas, in winter storms, the air lifting is driven by the horizontal temperature gradient induced by warm and cold fronts (i.e., frontal lifting). Orographic lifting was considered not to be relevant in Texas because of the overall low terrain relief, and dry air in the mountainous areas west of the 100° meridian. In general, storms caused by convective lifting are short in duration and spatially concentrated, although long and large summer storms caused by mesoscale convective complexes (MCC) can occasionally occur. Convective lifting is mostly an isotropic process in which direction is not a dominant variable. Storms produced by frontal lifting, on the other hand, tend to be longer in duration and larger in size. Frontal lifting is an anisotropic process in which the rainfall field assumes the orientation of the front that causes it (C. Schumacher, Texas A&M University – Department of Atmospheric Sciences, October 2005).

After subdividing the dataset into 40 subsets of records of the same duration and area (i.e., 5 × 8 combinations), sinusoidal regression equations of the ARF with the storm time of occurrence as predictor were developed. The sinusoidal equations had the form

$$ARF = a_1 - a_2 \sin\left(2\pi \frac{H - a_3}{H_{Max}}\right) \tag{2-2}$$

where H [hours] is the storm time of occurrence after January 1st 0:00 am, H_{Max} is the number of hours in a year (i.e., 8,760 hours), a_1 is the sinusoidal vertical offset, a_2 is the sinusoidal amplitude, and a_3 [hours] is the sinusoidal phase shift. For the 40 subsets, the values of a_3 ranged from 2,425 hours (April 11th) to 3,075 hours (May 8th) with an average value of 2,712 hours (April 23rd) and standard deviation of 215 hours (nine days). Accordingly, summer was defined from April 23rd to October 22nd and winter from October 23rd to April 22nd. Out of the 18,531 storms being studied, 15,825 (85%) took place in summer and 2,706 (15%) in winter. For the different durations, the percentage of summer storms ranged from 77% to 91%.

Likewise, based on climate conditions, Texas was subdivided into six regions according to USGS (1998) (Figure 2-3). These regions are aggregations of the TWDB (1967) climatic divisions. In this case, out of the 18,531 storms, 2,082 (11%) corresponded to region #1, 4,382 (24%) to region #2, 3,581 (19%) to region #3, 2,819 (15%) to region #4, 3,234 (17%) to region #5 and 2,433 (13%) to region #6. In all cases, the percentage of summer storms ranged from 81% to 89%.

2.2.4. Effect of the Storm Precipitation Depth on the ARFs

Once season and region were defined, the 148,248-record table was subdivided into 480 subsets such that each of them contained records with the same storm duration, area, season and region (i.e., $5 \times 8 \times 2 \times 6$ combinations).

In each of these subsets, high-leverage points and outliers were identified and removed before any type of analysis was conducted. Since the distribution of the number of storms with respect to depth was strongly positively skewed, a few storms with high rainfall depth would have had a considerable effect on the ARF-depth relation compared to that of the lower depth storms. To identify these high-leverage points, the distribution of the variable depth was normalized by an inverse transformation, and all points whose deviation from the mean divided by the standard deviation was less than -1.96 (i.e., on average the lower 2.5% values of the inverse of the depth) were flagged as high-leverage points. Statistical outliers, likewise, were identified using regression diagnostics that assess substantial changes in the fitted model (i.e., Cook's distance, DFBeta, DFFITS and FVARATIO) or in the discrepancy between estimated and observed values (i.e., SDRESID) caused by the inclusion or exclusion of individual points (Belsley et al. 1980). After removing high leverage points and outliers, for each of the subsets, regression equations of the form:

$$ARF = b_o + b_1 d \tag{2-3}$$

where d [L] is the precipitation depth, and b_0 (intercept) and b_1 (slope) [1/L] are regression coefficients, were determined. In case the ARFs did not depend on the depth with a level of significance of 0.05, the slope of the line b_1 was set equal to zero.

Finally, analyses of covariance (ANCOVA), between the subsets that had the same duration, area and region, but different season, or same duration, area and season, but different region, were run to determine if the ARF values for summer and winter or for different regions were statistically different with a level of significance of 0.05. If the datasets for summer and winter were found not statistically different, a single year-round equation was determined. Similarly, if the datasets for two regions were found not statistically different, they were merged into a single region and one equation was determined for both of them. Moreover, for three regions to be merged into a single region, the three two-way comparisons should find the regions not statistically different.

2.3. RESULTS

For given storm durations, areas, seasons, regions and precipitation depths, average ARF values are reported. Because of their close-to-normal distribution, these average values are likely to be exceeded half of the times. ARF values for other exceedance probabilities can be determined if necessary. The effect of the precipitation depth, season and region on the ARFs is discussed in this section. The distribution of storm ellipse aspect ratios and orientation angles is presented and contrasted for the different seasons and regions.

2.3.1. Effect of Depth, Season and Region

Tables 2-1 and 2-2 present the intercept and slope of the ARF-depth regression equations for summer storms. In Table 2-2, rows and values missing correspond to cases in which the effect of the depth is not relevant with a 0.05 level of significance (i.e., slope equal to zero). Similarly, Tables 2-3 and 2-4 present the equivalent values for winter storms. As with Table 2-2, in Table 2-4, rows and values missing correspond to cases in which the effect of the depth is not relevant with a 0.05 level of significance. In Tables 2-1 and 2-2, regions #4 and #5 are lumped together because the ARF values for summer storms in the two regions were not statistically different according to ANCOVA tests. No other two regions were found not statistically different for either summer or winter storms. ANCOVA tests also indicated that the season made a difference, with a level of significance of 0.05, in all regions except for a few cases in regions #1 and #6. In those cases in which region did not make a difference, the corresponding values in Tables 2-1 and 2-3 (intercept), and Tables 2-2 and 2-4 (slope) are equal.

Figure 2-6 shown previously, depicts ARF histograms for 1-hour storms and areas of around 737 km^2. It can be seen that the accounting of season, region and depth decreased the ARF variance from 0.026 for all storms (i.e., 7,479 storms), to 0.025 for summer storms (i.e., 6,486 storms), to 0.021 for summer storms in regions #4 and #5 (i.e., 938 storms), to 0.016 for summer storms in regions #4 and #5 and precipitation depth in the range from 53 mm to 69 mm (i.e., 100 storms). Note that the ARF variability is significant even after including season, region and depth in the analysis. Figures 2-7, 2-8 and 2-9 show ARF-area curves for 1-hour storms isolating the effect of depth, season and region, respectively.

Table 2-1. ARF equation intercept b0 for summer storms.

Region	Duration (hrs)	Area (km^2)							
		45	75	105	135	166	286	467	737
1	1	.82	.73	.68	.64	.60	.50	.41	.33
	3	.78	.67	.62	.58	.54	.46	.39	.32
	6	.78	.68	.64	.59	.55	.45	.38	.32
	12	.82	.76	.65	.60	.57	.55	.48	.35
	24	.83	.78	.73	.69	.65	.57	.50	.43
2	1	.79	.69	.64	.59	.56	.46	.39	.34
	3	.84	.77	.73	.69	.66	.57	.50	.43
	6	.85	.78	.70	.66	.63	.55	.52	.45
	12	.85	.78	.75	.71	.68	.61	.54	.47
	24	.86	.80	.76	.73	.71	.63	.57	.51
3	1	.78	.67	.61	.57	.54	.43	.34	.27
	3	.78	.69	.65	.61	.58	.52	.44	.39
	6	.78	.69	.64	.60	.57	.48	.41	.34
	12	.76	.67	.62	.57	.54	.44	.36	.31
	24	.80	.71	.67	.63	.59	.49	.42	.35
4 and 5	1	.76	.65	.59	.54	.50	.41	.33	.27
	3	.79	.70	.65	.61	.57	.48	.41	.35
	6	.80	.71	.66	.63	.59	.50	.43	.36
	12	.80	.72	.68	.64	.61	.52	.45	.39
	24	.81	.73	.68	.64	.61	.51	.44	.38
6	1	.79	.71	.66	.63	.59	.50	.47	.39
	3	.79	.72	.66	.62	.59	.52	.44	.39
	6	.80	.71	.67	.63	.60	.51	.44	.39
	12	.80	.71	.66	.62	.58	.50	.43	.38
	24	.79	.70	.64	.60	.57	.46	.39	.32

Table 2-2. ARF equation slope b1 for summer storms.

Region	Duration (hrs)	Area (km^2)							
		45	75	105	135	166	286	467	737
1	3	.0010	.0014	.0016	.0017	.0018	.0015	.0013	.0013
	6	.0008	.0012	.0012	.0014	.0015	.0017	.0016	.0015
	12			.0010	.0012	.0011			
2	1	.0006	.0011	.0013	.0013	.0013	.0012	.0008	
	6			.0005	.0007	.0007	.0007		
3	1	.0006	.0013	.0016	.0016	.0017	.0019	.0023	.0022
	3	.0007	.0010	.0011	.0011	.0011	.0009	.0010	.0007
	6	.0007	.0010	.0011	.0013	.0013	.0015	.0015	.0016
	12	.0009	.0012	.0014	.0016	.0017	.0020	.0022	.0022
	24	.0005	.0007	.0009	.0010	.0011	.0014	.0015	.0017
4 and 5	1	.0011	.0016	.0019	.0022	.0023	.0023	.0022	.0019
	3	.0006	.0010	.0011	.0011	.0012	.0013	.0012	.0011
	6	.0006	.0008	.0009	.0010	.0010	.0012	.0012	.0012
	12	.0005	.0007	.0008	.0009	.0009	.0011	.0011	.0011
	24	.0005	.0007	.0008	.0009	.0010	.0012	.0013	.0013
6	1	.0010	.0011	.0012	.0011	.0012	.0013		
	3	.0009	.0011	.0014	.0014	.0015	.0015	.0017	.0015
	6	.0009	.0011	.0013	.0014	.0015	.0018	.0018	.0017
	12	.0008	.0012	.0015	.0016	.0017	.0018	.0019	.0018
	24	.0008	.0012	.0015	.0016	.0017	.0021	.0022	.0024

Table 2-3. ARF equation intercept b0 for winter storms.

Region	Duration (hrs)	Area (km^2) 45	75	105	135	166	286	467	737
1	1	.82	.73	.68	.64	.64	.50	.45	.38
	3	.84	.78	.75	.70	.67	.59	.52	.46
	6	.83	.76	.72	.69	.67	.60	.54	.49
	12	.82	.76	.74	.70	.68	.61	.55	.50
	24	.83	.77	.73	.69	.71	.65	.60	.55
2	1	.86	.79	.75	.71	.68	.60	.52	.44
	3	.88	.82	.85	.84	.82	.76	.72	.67
	6	.89	.83	.80	.78	.76	.81	.77	.73
	12	.89	.83	.80	.77	.75	.69	.66	.59
	24	.90	.86	.81	.78	.78	.71	.66	.61
3	1	.88	.87	.83	.80	.78	.71	.64	.57
	3	.93	.90	.88	.86	.84	.76	.71	.64
	6	.77	.71	.67	.65	.63	.57	.53	.48
	12	.76	.70	.66	.64	.62	.57	.53	.50
	24	.74	.67	.64	.62	.60	.55	.51	.48
4	1	.85	.78	.74	.70	.67	.59	.51	.44
	3	.87	.81	.78	.75	.73	.66	.59	.53
	6	.87	.73	.69	.65	.63	.56	.47	.41
	12	.83	.74	.69	.66	.63	.56	.50	.41
	24	.80	.71	.66	.62	.59	.50	.43	.33
5	1	.86	.80	.76	.73	.70	.61	.53	.53
	3	.88	.83	.79	.76	.74	.67	.60	.54
	6	.88	.77	.80	.77	.75	.69	.62	.56
	12	.84	.79	.75	.72	.70	.70	.64	.58
	24	.88	.84	.81	.78	.76	.70	.64	.59
6	1	.84	.78	.74	.84	.81	.79	.77	.71
	3	.86	.80	.77	.74	.72	.66	.60	.54
	6	.87	.82	.79	.76	.74	.69	.64	.59
	12	.80	.71	.66	.62	.56	.47	.42	.38
	24	.88	.83	.80	.78	.77	.72	.67	.63

Table 2-4. ARF equation slope b1 for winter storms.

Region	Duration (hrs)	Area (km^2) 45	75	105	135	166	286	467	737
2	3			-.0013	-.0016	-.0018	-.0019	-.0025	-.0026
	6						-.0020	-.0023	-.0024
3	1	-.0013	-.0024	-.0024	-.0024	-.0024	-.0026	-.0025	-.0025
	3	-.0017	-.0022	-.0024	-.0025	-.0025	-.0024	-.0023	-.0021
4	6		.0012	.0014	.0015	.0016	.0017	.0018	.0019
	12	.0007	.0011	.0013	.0015	.0016	.0018	.0017	.0021
	24	.0011	.0016	.0018	.0020	.0022	.0025	.0027	.0032
5	1								-.0017
	6		.0008						
	12	.0005	.0005	.0005	.0006	.0006			
6	1				-.0035	-.0035	-.0055	-.0071	-.0074
	12	.0008	.0012	.0015	.0016	.0025	.0029	.0029	.0029

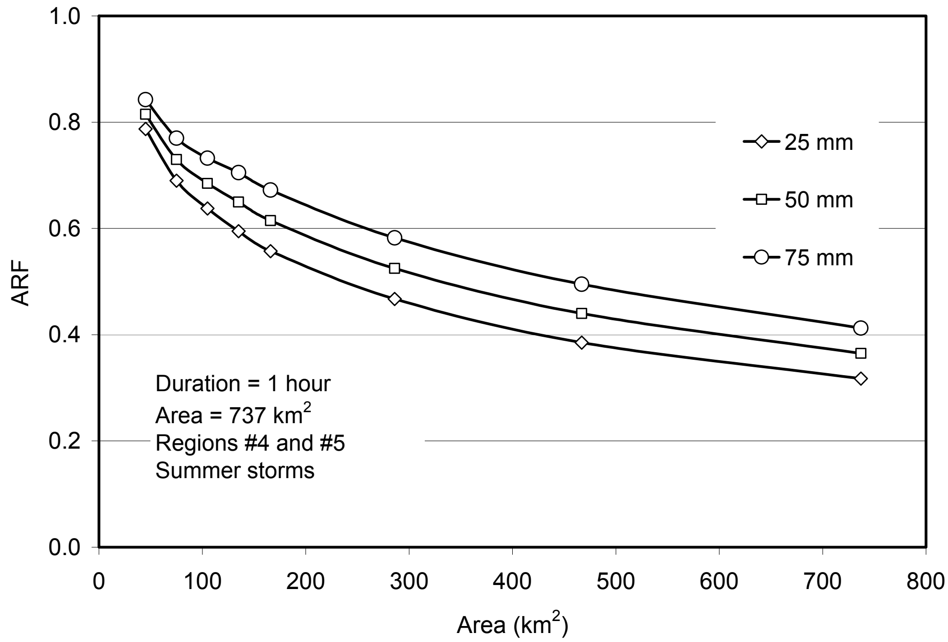


Figure 2-7. ARF vs. area curves for different precipitation depths.

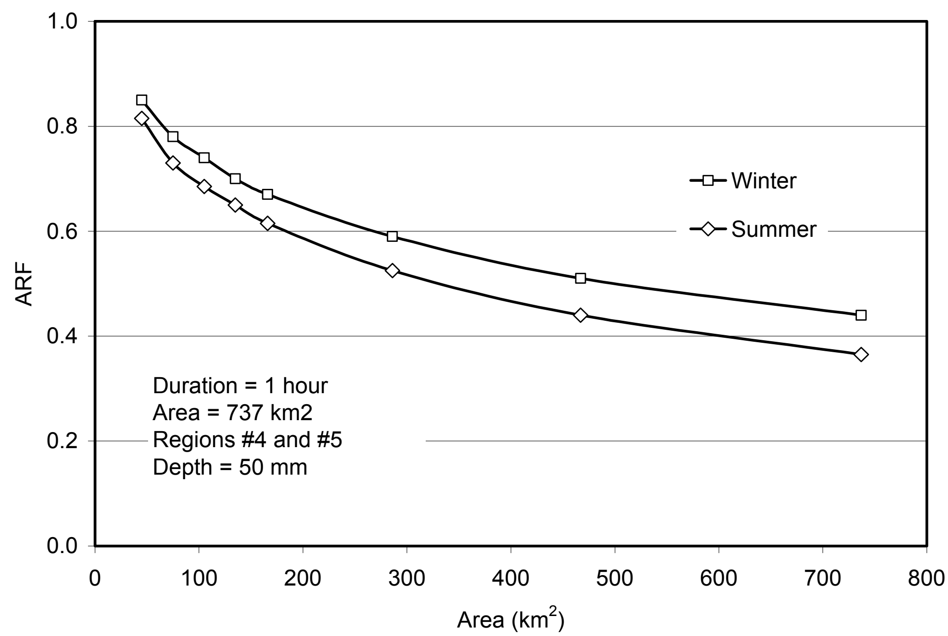


Figure 2-8. ARF vs. area curves for summer and winter storms.

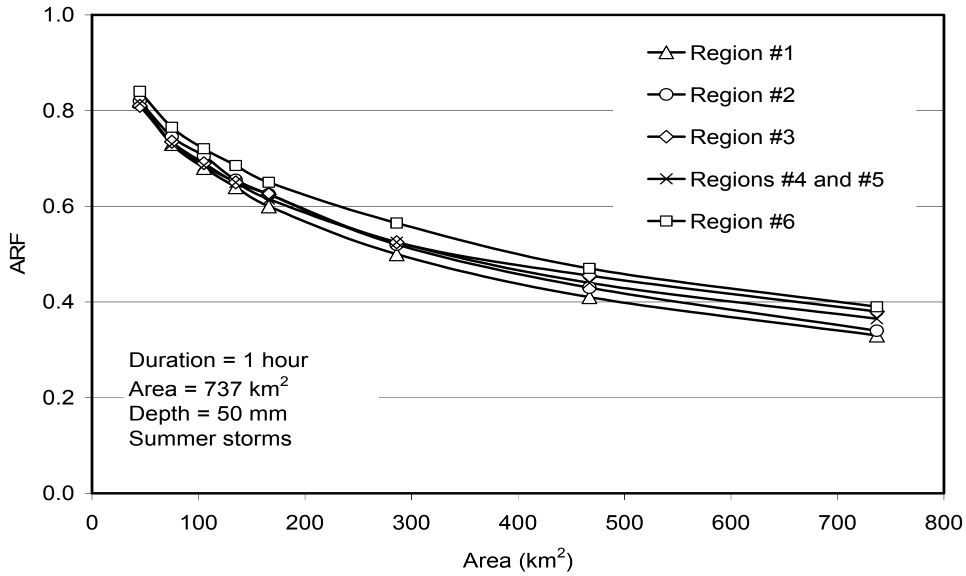


Figure 2-9. ARF vs. area curves for different regions.

For the dataset used, it was found that season and region were variables that had to be considered in the analysis. With respect to depth, it was observed that it was more relevant for summer than for winter storms. In fact, in some specific cases of winter storms, it was even found that depth had a negative effect (i.e., ARFs decreased with depth) (see Table 2-4), which could have been caused by the relatively small size of the winter dataset. It should also be kept in mind that the range of depth values used to derive the ARF-depth regression equations did not include high precipitation depths associated with low frequency storms (see Table 2-5), and that the application of these equations should be limited to the same range of values. In fact, use of these equations outside its range of applicability may result in ARFs greater than one. A significantly longer period of record (something not available yet in radar-based estimated rainfall data) would be necessary to include low frequency values in the dataset.

Table 2-5. Range (mm) of the precipitation depths.

Region	Duration (hrs)	Lower Limit	Upper Limit	
			Summer	Winter
1	1	20	62	89
	3	25	94	110
	6	30	111	136
	12	35	122	136
	24	40	130	138
2	1	20	67	57
	3	25	105	94
	6	30	120	114
	12	35	132	114
	24	40	143	134
3	1	20	75	75
	3	25	120	130
	6	30	145	163
	12	35	170	192
	24	40	190	243
4	1	20	83	69
	3	25	136	104
	6	30	162	144
	12	35	180	145
	24	40	204	168
5	1	20	83	84
	3	25	136	141
	6	30	162	177
	12	35	180	220
	24	40	204	245
6	1	20	62	60
	3	25	109	103
	6	30	143	114
	12	35	160	126
	24	40	189	143

2.3.2. Aspect and Angle of the Optimum Ellipses

For the data available, it was possible to determine predominant storm ellipse aspect ratios and orientation angles, and their seasonal and regional distribution. Normalized histograms of aspect for storm durations of 1 hour are included in Figure 2-10. Similar histograms were prepared for durations of 3, 6, 12 and 24 hours. Based on Figure 2-10, it can be said that 31% of the summer storms and 28% of the winter storms have aspect values lower than 2, which corresponds to almost-circular shapes, while the remaining 69% and 72% have a clearly elongated elliptical shape. For aspect values greater than 2, it was also observed that the frequency decreased with the aspect, but increased again for the bin of 4.5 to 5. This abnormal increase was caused by the setting of 5 as the maximum aspect ratio in the determination of the optimum ellipses. Overall, patterns similar to those shown in Figure 2-10 were observed for other storm durations.

Similarly, normalized histograms of the orientation angle for storm durations of 1 hour and areas of around 737 km^2 are presented in Figure 2-11 for those ellipses with aspect ratio greater than 2. The aspect threshold of 2 was used because the angle was considered trivial for almost-circular ellipses. Likewise, the histograms include only 737-km^2 ellipses with the understanding that large ellipses capture better the orientation pattern of the rainfall field than the small ones. Again, similar histograms were prepared for durations of 3, 6, 12 and 24 hours, but were not included because of space limitations. In these histograms, an angle of 0° referred to an ellipse whose longer axis is parallel to the east direction, and angles were measured counterclockwise.

It can be seen that winter storms (white bars) have a strong preference for southwest-northeast orientations (i.e., angles lower than 90°) with around 80% of them having these orientations. This preference, however, is weaker for summer storms (black bars) in which only around 60% have southwest-northeast orientations. These ellipse aspect and angle histograms confirm the anisotropic character of the precipitation fields and the need of accounting for it in the analysis.

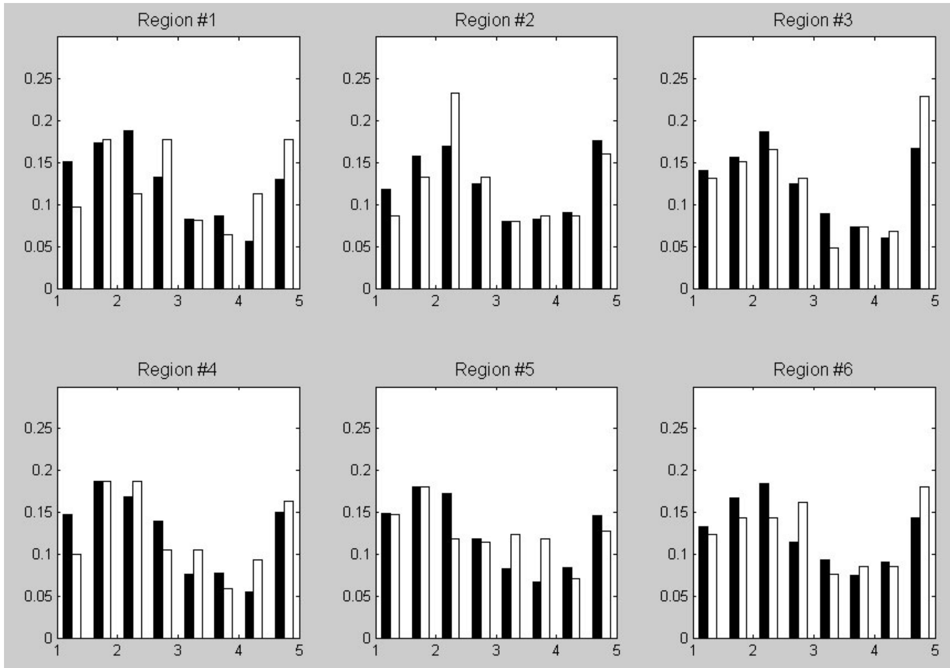


Figure 2-10. Aspect histogram for 1-hour storms. Black represents summer storms and white winter storms.

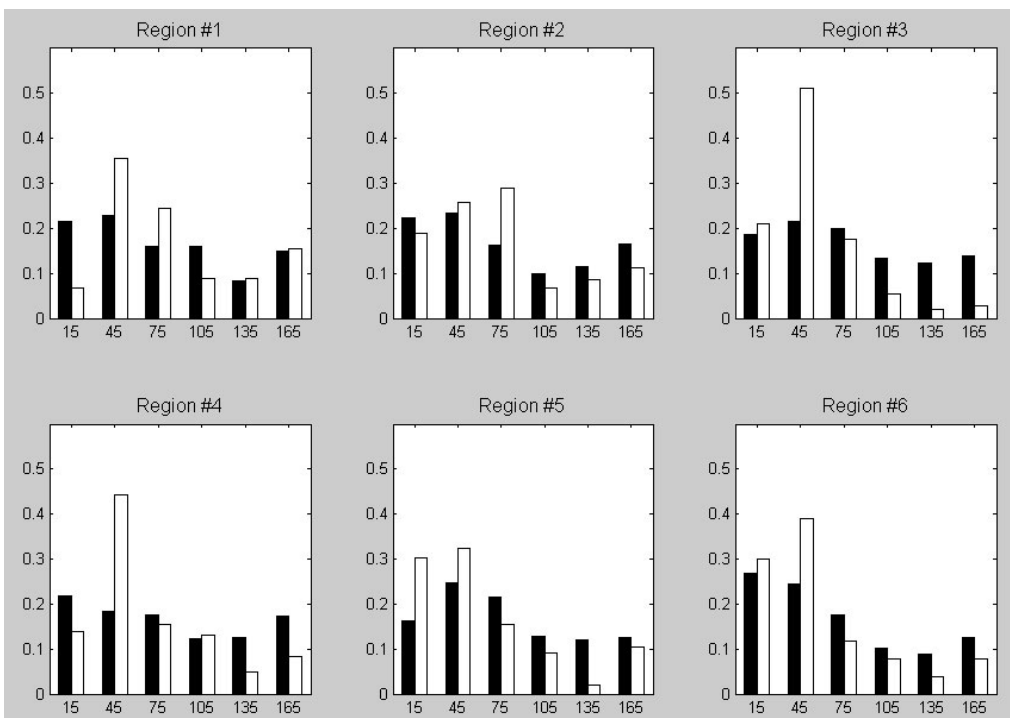


Figure 2-11. Orientation angle histogram for 1-hour storms. Black represents summer storms and white winter storms.

2.4. SUMMARY AND CONCLUSIONS

Average precipitation areal reduction factors (ARFs) for the 685,000 km^2 of Texas were calculated using NEXRAD rainfall estimates of years 2003 and 2004. The study was based on 18,531 selected snapshots of storms of duration of 1, 3, 6, 12 and 24 hours. Two seasons (summer from April 23rd to October 22nd and winter form October 23rd to April 22nd) and six regions (Figure 2-3) were used in the analysis. It was found that, in addition to the storm duration and area, other factors such as the season and region in which the storm took place, and the precipitation depth had a statistically significant effect on the ARFs. The effect of the precipitation depth on the ARFs was found to be stronger in summer than winter storms. However, because of the relatively narrow depth range available in the database, which resulted from the use of only two years of precipitation data, a complete understanding of the effect of depth on the ARFs could not be determined.

For 1-hour duration storms, it was observed that 31% of the summer storms and 28% of the winter storms have aspect values lower than 2, which were considered “almost circular”. The remaining 69% and 72% for summer and winter storms, respectively, had a more elongated elliptical shape. It was also observed that 60% of the summer storms had southwest-northeast orientations (i.e., somewhat parallel to the Texas Gulf coast), while for winter storms this percentage was 80%. These results are in line with the fact that frontal lifting, which is more frequent in winter than in summer, tends to generate storms parallel to the front that causes them. Convective lifting, on the other hand, tends to generate storms with no defined orientation preference. Similar results were observed for other durations. Note that, the greater number of observations in summer than in winter makes the summer results statistically more robust.

In general, lower ARF values were observed compared to previous studies (i.e., Weather Bureau 1957, Asquith and Famiglietti 2000). Possible explanations of these differences include:

(1) our ARFs relate concurrent area to point precipitation, rather than annual mean values; (2) our ARFs account for the anisotropy of the precipitation field; (3) our ARFs were based on distributed NEXRAD precipitation estimates, rather than on point gauge precipitation records; (4) our ARFs were based on a two-year dataset, rather than long precipitation records; and (5) individual-storm ARFs have significant variability. Overall, variability in the ARF values was observed and, even though part of this variability could be explained by seasonality, regionality and precipitation depth (associated to the storm frequency), the uniqueness of each storm event appears to be an important cause of it.

3. HOURLY DISAGGREGATION OF DAILY RAINFALL IN TEXAS USING MEASURED HOURLY PRECIPITATION AT OTHER LOCATIONS

A method to disaggregate daily rainfall into hourly precipitation is evaluated across Texas. Based on Socolofsky et al. (2001), the method chooses representative storm intensity patterns from measured hourly databases to generate the synthetic data using a single parameter for the smallest expected one-hour event. The model is applied across Texas using historic hourly precipitation data; performance is evaluated by the model's ability to reproduce hourly rainfall statistics. Based on a cluster analysis to determine which precipitation databases should be used for disaggregation, no trends in space or among gauge characteristics (e.g. period of record, precipitation statistics) were identified. As a result, a Texas state database containing all the measured hourly data in Texas is proposed for use by disaggregation. The state database is verified for a selection of gauges and performed as well at a given station as using that station's measured rainfall for the disaggregation. The method is further applied to estimate intensity–duration curves which show that the method matches the majority of storm intensities needed to track soil moisture and diverges by less than 17% for the extreme runoff-generating events.

3.1. INTRODUCTION

Partitioning rainfall into surface and subsurface pathways is a fundamental goal of watershed modeling and provides the critical forcing mechanism to many watershed processes (Chow 1964; Eagleson 1970; Phillip and Wayne 1992). To perform the partitioning, local soil moisture plays an important role and is greatly affected by the intermittency and intensity of rainfall. As a result, continuous simulation models for watershed hydrology typically require sub-daily rainfall data as input. While hourly data, for example, may be available for recent years or a few gauges only, continuous simulation models require many years of data for calibration and verification. In the United States, historic daily rainfall records are nearly three times as dense as hourly gauges (Bonner 1998), yielding a strong motivation to disaggregate the daily data into hourly time series for watershed modeling. The disaggregated hourly time series are inherently synthetic, and can be drawn from either rainfall simulation models or measured rainfall databases. Socolofsky et al. (2001) introduced a method using the latter approach in which an hourly storm database is stochastically sampled to reproduce the measured daily total. Though this method appears fruitful, it has only been demonstrated for a few gauges in Massachusetts. Here, we evaluate its performance to generate synthetic time series of hourly rainfall from daily totals across Texas. Applying this method over a large region like Texas provides significant guidance for applications in other states or countries.

To effectively predict runoff, the input rainfall to hydrologic models must accurately capture rainfall intensity and storm intermittency (Donigian et al. 1984; Marani et al. 1997; Margulis and Entekhabi 2001; William 1981; William and Peter 1981). For real-time flood forecasting, the time of the storm must be known, and the input rainfall data must be measured directly within the watershed. Conversely, for comparison to daily totals at the watershed outlet, as is often done for long-term simulations, the time of the storm becomes less important as the time of concentration of the watershed approaches an order of one day or more. In such cases, the storm intensity structure is more important than the time of rainfall, and in this article we assume that for the disaggregated data to be effective they must only match the rainfall statistics and measured daily totals at the disaggregation site and not the actual time of rainfall.

Two broad categories of methods are available to obtain rainfall with the correct statistics: methods based on stochastic simulation of precipitation and deterministic methods based on direct measurement of daily precipitation depth from an off-site gauge. In either case, these rainfall data must be further conditioned to match the measured daily rainfall to achieve disaggregation.

Stochastic simulation time series are generally obtained by continuous-simulation stochastic rainfall models or through fractal random cascade models. Other stochastic approaches attempt to fit probability distributions to rainfall characteristics and use these distributions to infer hourly intensity patterns (Econopouly et al. 1990; Hershenhorn and Woolhiser 1987; Woolhiser and Osborn 1985). Most stochastic rainfall models are based on the Bartlett-Lewis or Newman-Scott rectangular pulses models (Cowpertwait et al. 1996a; Cowpertwait et al. 1996b; Islam et al. 1990; Onof and Wheater 1993; Onof and Wheater 1994; Rodriguez-Iturbe et al. 1987; 1988). These models utilize five to seven parameters to simulate rainfall as a Poisson arrival process of storms and storm cells. The advantage of these synthetic data is that they are of a continuous nature, able to be re-sampled at any desired aggregation level. Various fractal random cascade models make use of the scale invariance of the rainfall process over a range of time scales (Glasbey et al. 1995; Gupta and Waymire 1993; Hingray and Haha 2005; Molnar and Burlando 2005; Olsson 1998; Olsson and Berndtsson 1998; Ormsbee 1989; Schertzer and Lovejoy 1987; Veneziano et al. 1996). These models capture the variability both between storms and within storms and require the fitting of scaling laws to apportion rainfall at different scales.

For the stochastic simulation methods, two types of disaggregation are identified. In the first case, called downscaling (Koutsoyiannis and Onof 2001), the model parameters are estimated for measured data at one time scale (e.g. daily) and used to obtain simulated rainfall at a lower (e.g. hourly) time scale (see e.g. Bo et al. 1994). In the second case, which we call direct disaggregation, the synthetic data are used to reproduce measured daily rainfall totals by either sampling or conditioning the simulated rainfall (Glasbey et al. 1995; Gyasi-Agyei 2005; Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003). The Bartlett-Lewis rectangular pulses models are especially suited to direct disaggregation because they can be run in an iterative approach (see for example Gyasi-Agyei 2005; Koutsoyiannis and Onof 2001; Koutsoyiannis et al. 2003), but one limitation of the method is its tendency to over-estimate extreme events (Gyasi-Agyei 2005). The fractal cascade models are also appropriate for direct disaggregation using either a canonical cascade, where the rainfall volume is preserved in the mean, or a micro-canonical cascade, where the total rainfall volume is preserved exactly at each cascade level (Olsson 1998). While efficient at rainfall scale transformation, each of these approaches uses multiple parameters to describe scaling laws, branching weights and probability distributions.

As an alternative to stochastic methods, deterministic rainfall disaggregation utilizes known weather patterns or measurements from nearby gauges. Gutierrez-Magness and McCuen (2004) evaluated four types of deterministic rainfall disaggregation methods for 74 gauges near the Chesapeake Bay watershed in the United States. These methods included uniform distribution of daily rainfall over 24 hours, use of weather pattern methods utilizing meteorological distribution patterns, direct transfer of hourly data from one gauge to another location, and scaling transfer where the non-dimensional intensity pattern at one gauge is used to disaggregate the measured daily rainfall at another gauge (Gutierrez-Magness and McCuen 2004). Because the exact time of rainfall is unknown for disaggregation, none of the methods performed well when directly compared to the measured data (Gutierrez-Magness and McCuen 2004). Although deterministic approaches are based on measured hourly rainfall, they are not expected to predict well the intensity structure of disaggregated storms because of their inherent need to adjust the scale of measured precipitation.

The approach we apply here captures better the intensity structure of storms than do deterministic methods and uses measured data rather than simulated stochastic precipitation. Described in the Data and Disaggregation Methods section below and in more detail in Socolofsky et al. (2001), the method utilizes a one-parameter stochastic selection algorithm to simulate daily totals using measured storm intensity patterns at an off-site hourly recording database and has not been applied outside of a limited study in Massachusetts. The use of measured data eliminates the need for a rainfall simulation model, and the selection process ensures that unscaled, representative storms are used in the disaggregation. The single parameter in the disaggregation method represents the smallest expected one-hour precipitation amount and in general must be calibrated to measured data. The Data and Disaggregation Methods section also describes the available data and an automated calibration scheme for the smallest storm parameter based on Bayesian estimation. The Results and Discussion section describes the method performance. By applying spatial analysis, we show that the identification of the storm databases used for disaggregation does not show spatial or climatic patterns. This fact is exploited in the Application section which demonstrates the performance of the disaggregation scheme for a set of verification gauges using a single, lumped state of Texas rainfall database.

3.2. DATA AND DISAGGREGATION METHODS

In Texas, over 2100 rain gauges are distributed over 695,670 km^2. Of the over 2100 gauges, only 646 are hourly recording, and these have variable periods of record from a few months to over 50 years. A valid daily rainfall disaggregation tool, therefore, will significantly increase the coverage of hourly precipitation data available for long-term hydrologic simulation, calibration, and verification.

3.2.1. Rainfall Data

Hourly precipitation data for Texas were obtained from a commercially available database (EarthInfo 2001) based on data set 3240 of the National Climate Data Center (NCDC, 2003) . For Texas, the data included a total of 646 rain gauges. Out of these 646 gauges, 115 did not include location information (i.e., longitude and latitude) and were excluded from our analysis, leaving the 531 hourly gauges shown in Figure 3-1. The data coverage was from 1900 to 2001, with an average record start time of 1952 and an average period of record of 18 years. The longest period of record was 61 years and occurred at 36 gauges. Overall, average annual precipitation in Texas increases in the West – East direction from 230 mm to 1500 mm (TWDB 2006).

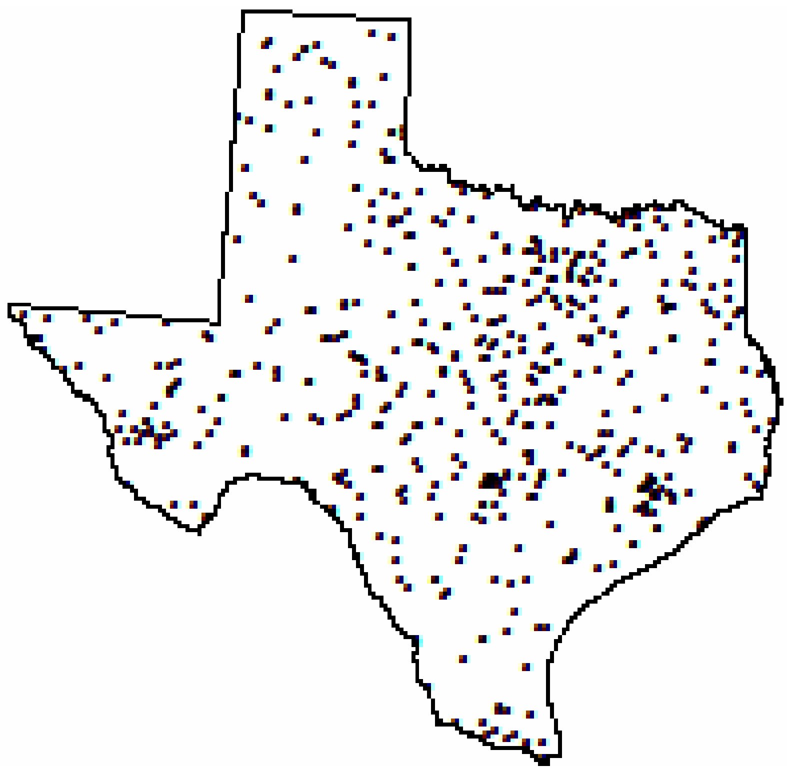


Figure 3-1. Locations of the 531 NCDC rain gauges in Texas with hourly data.

3.2.2. Rainfall Disaggregation Scheme

For disaggregation, we apply the stochastic storm selection method (Socolofsky et al. 2001). The method utilizes a database of measured storm events to disaggregate the measured daily rainfall; the storm database may be an off-site gauge or a collection of gauges.

The disaggregation method comprises the following main steps. For each month of the year, a database of storm events (defined as sequences of uninterrupted hourly rainfall) is created from the hourly recorded data. Each of the storms for a given month are ranked by their total precipitation volume to obtain the cumulative density function (CDF) for storm volume for the month. DT defines the total depth of daily rainfall measured at the aggregated gauge. A search algorithm finds the limit a in the monthly storm CDF such that all storms with a probability less than a have a depth less than DT. A uniform random number u is then selected between 0 and a. The storm in the database that falls closets to u is selected and becomes parts of the intensity pattern for the disaggregated data series. The total rainfall depth is then updated by DT = DT – D(u), and the selection sequence is repeated as long as DT is above a threshold value ε.

The single parameter in the disaggregation scheme ε is physically related to the smallest expected one-hour storm event. The ε parameter is needed to prevent the method from selecting a large number of trace events when the remaining daily rainfall is small. When DT is less than ε, the method stops selecting storms from the measured database, and the remaining rainfall is simulated from an exponential distribution with mean depth equal to the remaining DT.

As in any disaggregation method, the start times for the storms are unknown and several options can be used to estimate them. In this dissertation, we use the measured times in the storm database. Previously, Socolofsky et al. (2001) chose start times from a uniform probability distribution. Our results do not show sensitivity to either case. We use the measured start times here because they are expected to more closely match the distribution of rainfall throughout the day. The start time for the final, one-hour storm due to ε is selected from a uniform probability distribution between 1 and 24 hours. Using these methods, storms inevitably overlap, and overlapping intensity patterns are summed to obtain the final hyetograph for the day.

3.2.3. Performance Assessment

To evaluate the performance of the stochastic storm selection method, all of the hourly recording rainfall gauges in Texas are used. From this data, we perform two types of disaggregation. First, auto-disaggregation uses the measured hourly data at one gauge to disaggregated that gauge's daily totals. During auto-disaggregation, all the measured hourly storms at a given daily gauge are present in the database so that the daily rainfall could potentially be disaggregated by the actual hourly rainfall measured at the gauge, but this is not guaranteed or expected. Auto-disaggregation is used both to calibrate the value of ε in the model and to obtain a guideline for the best level of performance expected by the stochastic storm selection method. Second, general disaggregation uses a database of off-site measured storms to disaggregate a gauge's daily totals. In this case, the off-site data could be another gauge or collection of gauges. The daily gauge is obtained by aggregating the hourly data at one of the Texas gauges. Hence, in both types of disaggregation, the daily data are aggregates of an hourly gauge so that the success of the disaggregation can be compared to the measured hourly data.

As pointed out by Gutierrez-Magness and McCuen (2004), disaggregated time series cannot be compared to the measured data on an hour-by-hour basis because of the uncertainty in the start times of storms in the disaggregated data. As a result, we take the statistical approach suggested by Socolofsky et al. (2001) to evaluate performance. They identified four important measures that should be matched by the disaggregated data: conservation of mass, the probability of zero rainfall, the variance of hourly rainfall, and the lag-1 hour autocorrelation coefficient. Each of these statistics is calculated on a monthly basis for the measured and the disaggregated data. For calibration, the parameter ε is adjusted each month using auto-disaggregation until a best compromise is achieved between the measured and modeled values for these four statistics. For validation, the performance of general disaggregation is compared to the measured statistics and to the auto-disaggregation data statistics. Using these techniques, we calculate the spatial distribution of ε in Texas for each month and develop guidelines for selecting disaggregation databases for use in general disaggregation.

3.2.4. Automated Calibration

To apply the disaggregation method to all 531 of the hourly rainfall gauges in Texas, we use an automated calibration approach for the threshold storm size parameter ε. We utilize our knowledge of reasonable values for a small, one-hour storm event by applying a Bayesian approach for parameter estimation (Schweppe 1973; Socolofsky and Adams 2003). In this approach, we minimize the objective function

$$J(\varepsilon) = [z - f(\varepsilon)]^T C_v^{-1} [z - f(\varepsilon)] - [\varepsilon - \bar{\varepsilon}]^T C_z^{-1} [\varepsilon - \bar{\varepsilon}] \qquad (3\text{-}1)$$

for each month, where the first term is the standard weighted least-squares estimator and the second term is the Bayesian regularization term which penalizes values of ε that are far from an expected value $\bar{\varepsilon}$. z is a vector of measured statistics, $f(\varepsilon)$ is a vector of statistics for the disaggregated data for a given value of ε, and the prior estimate $\bar{\varepsilon}$ is taken as the monthly value of ε from the Massachusetts watershed application (Socolofsky et al. 2001). C_υ^{-1} is a weighting matrix of covariances for the statistics used in the optimization (we used the four parameters described in the Performance Assessment section above with equal weighting) and C_ε^{-1} is a weighting matrix for the variance in ε, taken as 100 mm^2. By adjusting C_ε^{-1}, we give greater or lesser weight to our expected values $\bar{\varepsilon}$. The method naturally converged on positive values of ε so that Lagrange multipliers were not needed. Because $f(\varepsilon)$ is nonlinear, the minimization is achieved by Taylor expansion about the prior estimate $\bar{\varepsilon}$ to obtain an iterative secant method solution to the nonlinear problem.

3.3. RESULTS AND DISCUSSION

3.3.1. Distribution of Smallest-Storm Calibration Parameter ε across Texas

For each month, the smallest-storm parameter ε for each of the 531 gauges was estimated such that it minimized the squared difference between the statistics of the actual precipitation time series (i.e. measured statistics) and the statistics of the time series obtained by auto-disaggregation (i.e. auto-disaggregation statistics). We assume that the auto-disaggregation statistics match the true statistics better than the statistics resulting from disaggregation using any other gauge as the storms database. Any mismatch between the true statistics and the auto-disaggregation statistics constitutes an inherent error of the method, which is not caused by climatic or geographic differences between the disaggregated and disaggregating gauges. As a result, this mismatch provides a guideline for the expected error in the general disaggregation process.

Once the ε values are estimated for each month and gauge point, 12 interpolated surfaces were developed, so that ε values could be estimated for each month and at every location in Texas. The ordinary kriging method (Johnson et al. 2001) was used to create the surfaces, which required normalizing the distribution of the ε values and removing the outliers. Box-Cox transformations (Box and Cox 1964) were performed to normalize the distribution, and values more than three standard deviations from the mean were flagged as outliers and removed. The parameter of the Box-Cox transformation was set interactively for each month such that it minimized the kurtosis and the difference between the mean and the median of the distribution of ε values. Figure 3-2 presents the interpolated ε surfaces for February and August. Note that the variability of the ε values depends more on the month of the year than on the location. This seasonal variation was expected because of the different mechanisms that cause storms in summer and winter. Generally speaking, winter storms in Texas are associated with frontal air lifting, while summer storms are associated with convective air lifting. The ε values, according to the interpolated surfaces, range from 0.04 mm to 24.26 mm for February with a mean of 4.44 mm, standard deviation of 3.36 mm and skewness of 1.78 mm; and from 3.71 mm to 33.27 mm for August with mean of 12.5 mm, standard deviation of 4.06 mm and skewness of 1.37 mm. By cross validating the observed and paired interpolated ε values, the mean error was -0.04 mm and the root mean square error 3.40 mm for February, and 0.01 mm and 4.11 mm for August.

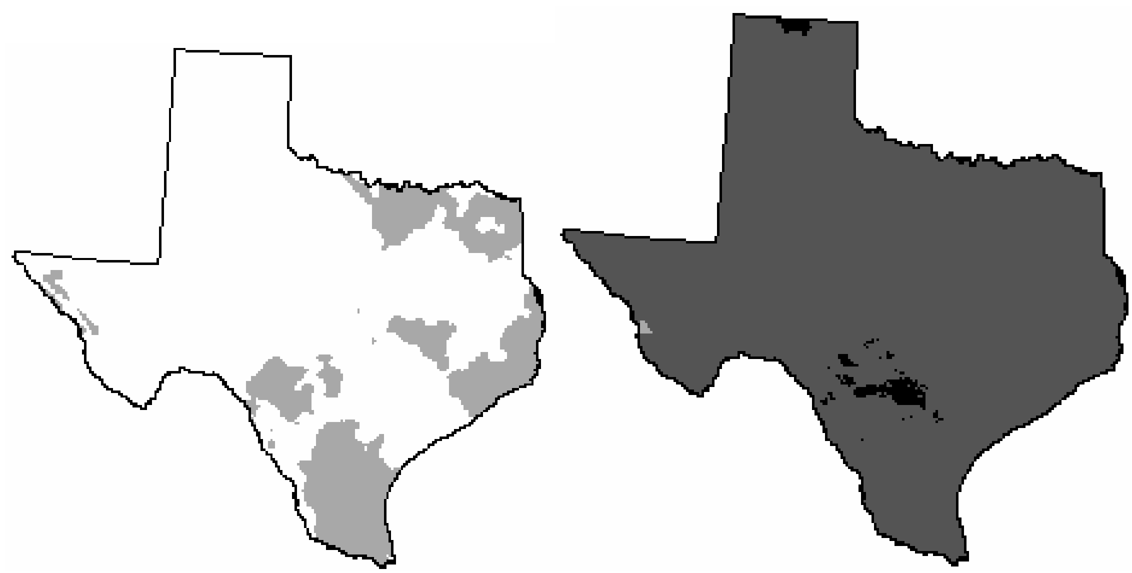


Figure 3-2. Surfaces for ε for the months of February (left) and August (right). Legend – white: 0 mm – 5 mm, gray: 5 mm – 10 mm, dark gray: 10 mm – 25 mm, black: 15 mm – 20 mm.

The relatively weak spatial variability in ε for a given month suggests it is possible to use a constant value of ε across the state of Texas for each month. Figure 3-3 shows the average ε values with one standard deviation error bars. Lower values of ε occur in the cooler months, when frontal storms are more likely and the precipitation can have long periods of light rain. Higher values of ε occur in the warmer months, when convective thunder showers are more likely, leading to short, intense storms. Although these monthly averages could have been used in our performance evaluation, the remaining calculations were obtained by using the ε values interpolated from the ε surfaces.

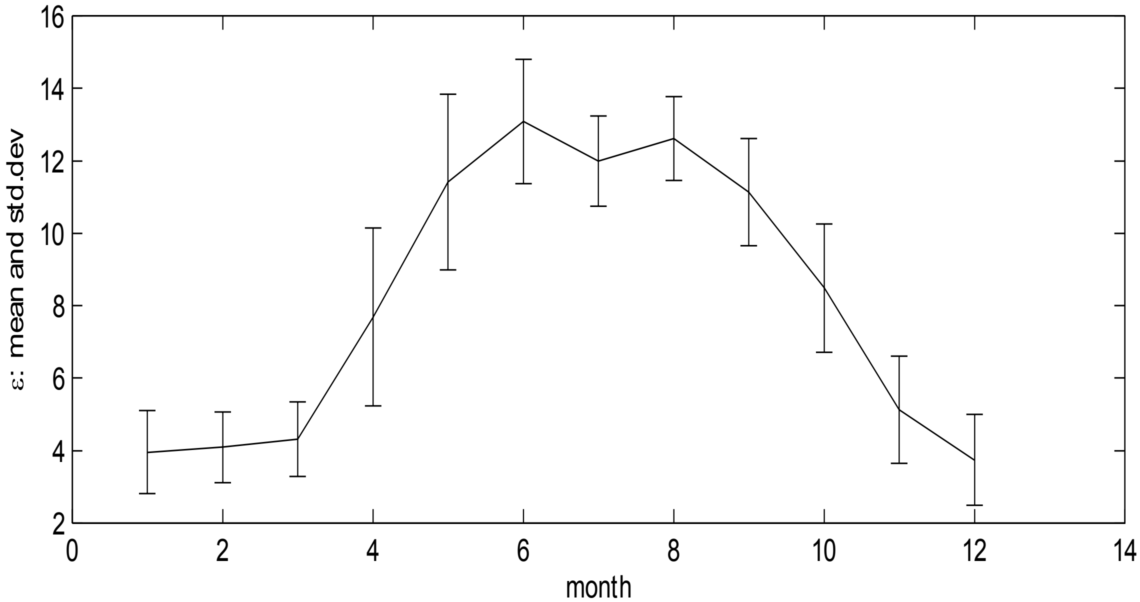


Figure 3-3. Spatial average of ε as a function of month with plus or minus one standard deviation.

3.3.2. Auto-disaggregation Performance

To evaluate the auto-disaggregation performance, the statistics of the auto-disaggregation time series at each gauge are compared with the true statistics. In this comparison, only 323 out of the 531 gauges were considered. The other 208 gauges consisted of 185 gauges that had periods of record less than five years, which was deemed too short for the data to have statistical value, and 23 gauges were reserved for model verification.

As suggested by Willmott (1982), the following measures of performance were used: mean absolute error (MAE), relative root mean square error (RMSE), linear regression (intercept

b_0 and slope b_1), coefficient of determination (r^2), and index of agreement (d). Each of these measures is well-known, except perhaps the index of agreement, which is defined as

$$d = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} \left(| P_i - \overline{O} | + | O_i - \overline{O} | \right)^2} \tag{3-2}$$

where P_i is the auto-disaggregation statistic of gauge i, O_i is the true statistic of gauge i, n is the number of gauges (i.e., n = 323), $\overline{P}$ is the average of the auto-disaggregation statistics, and $\overline{O}$ is the average of the true statistic. These metrics were calculated for each statistic and month and are presented for February and August in Tables 3-1 and 3-2. Note that it is the statistics of the time series that are being compared and not the time series themselves; thus, well-known efficiency metrics, such as the Nash-Sutcliffe coefficient (Nash and Sutcliffe 1970) and others, were not considered for performance evaluation.

Each of the metrics used relate to different aspects of the model performance. A MAE or RMSE value of zero corresponds to a perfect match between the two datasets. The linear regression coefficients help to evaluate the model bias; values for b_0 of zero and b_1 of one correspond to zero bias between the observations and predictions. The r^2 values quantify the scatter about the regression line. An r^2 value of one corresponds to perfect agreement with the regression line, but r^2 is insensitive to systematic model bias, because it does not account for line slopes different from one and intercepts different from zero (Willmott 1982). To overcome this shortcoming, the index d was designed to overcome the insensitivity of r^2 to additive and proportional differences between the model simulations and observations (Legates and McCabe 1999). A d value of one corresponds to perfect agreement.

In Tables 3-1 and 3-2 and Figure 3-4, it can be seen that the zero rainfall probabilities (P_0) obtained by auto-disaggregation are close to the true values. The variances (σ^2) show some bias toward systematic under-prediction, and this tendency is more evident as the true variance increases. This fact was particularly strong in the month of February. The lag one-hour autocorrelation coefficients (ρ_1) show the greatest scatter and tend to be over-predicted for lower values and under-predicted for higher values. Because the data used for the auto-disaggregation are at the same location as the disaggregated gauge, these performance metrics specify the best performance expected by the stochastic storm selection method.

Table 3-1. Calculated statistics for the observed and auto-disaggregated data.

Month	Statistic	n	$\bar{P}$	$\bar{O}$	s_p	s_o
February	P_0	323	0.9759	0.9750	0.0164	0.0165
	σ^2	323	0.0004	0.0009	0.0004	0.0022
	ρ_1	323	0.2339	0.2792	0.1154	0.1580
August	P_0	323	0.9721	0.9733	0.0171	0.0160
	σ^2	323	0.0005	0.0008	0.0004	0.0007
	ρ_1	323	0.2520	0.3154	0.1040	0.1411

Table 3-2. Performance measures between the observed and auto-disaggregated data.

Month	Statistic	MAE	b_0	b_1	RMSE	r^2	d
February	P_0	0.0025	0.0281	0.9721	0.0045	0.9569	0.9882
	σ^2	0.0005	0.0003	0.1074	0.0023	0.2939	0.3419
	ρ_1	0.1156	0.1554	0.2812	0.2285	0.1481	0.6318
August	P_0	0.0031	-0.0340	1.0338	0.0055	0.9384	0.9819
	σ^2	0.0003	0.0001	0.4588	0.0007	0.7544	0.7810
	ρ_1	0.1174	0.1604	0.2903	0.2139	0.1550	0.6116

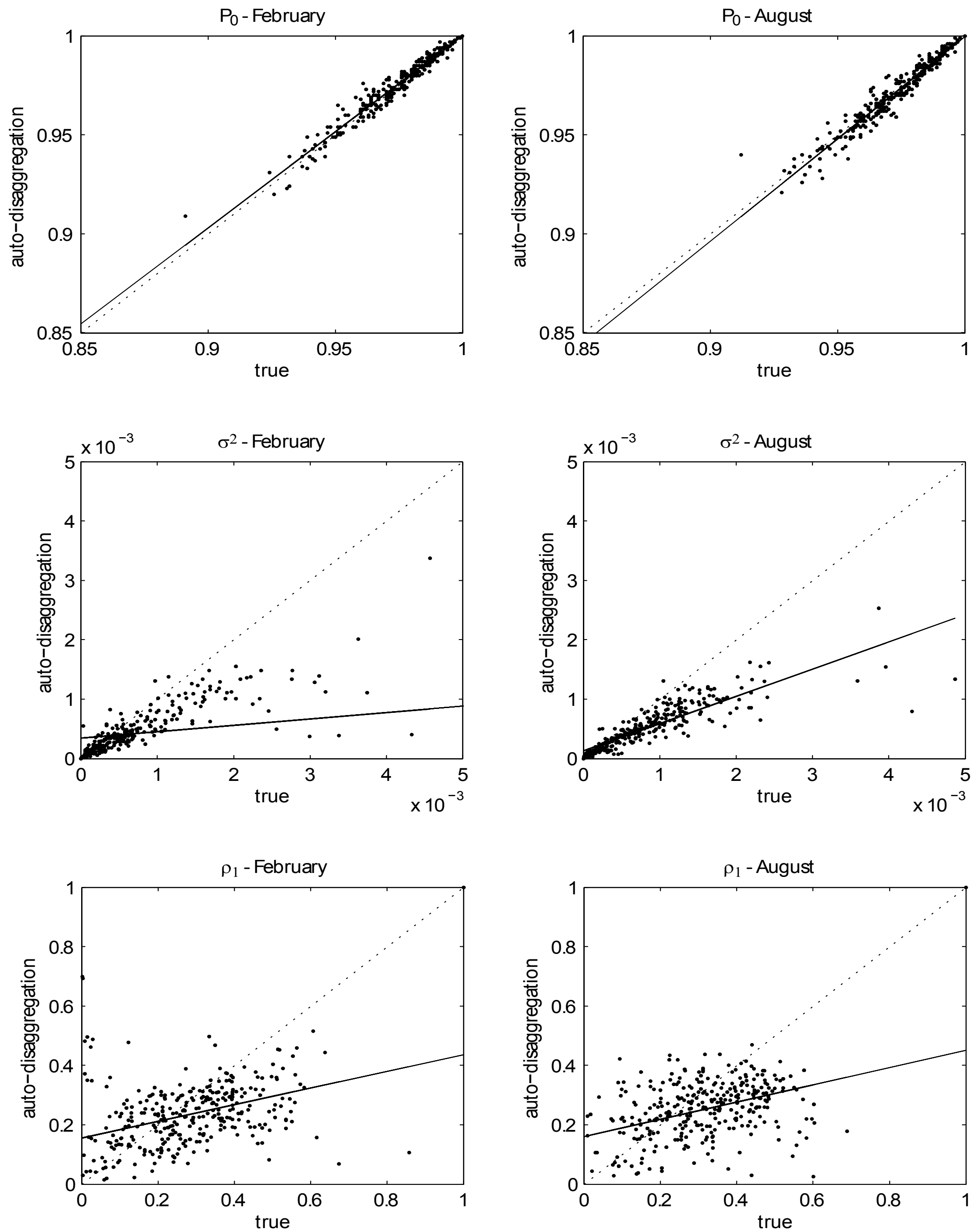


Figure 3-4. Comparison between auto-disaggregation and true statistics. The solid line represents a linear regression model. The dashed line represents the 1:1 line (a perfect model).

Gauge performance trends

For general disaggregation, the question arises: which hourly storm database should be used to disaggregate a given gauge location? While nearby gauges may be expected to perform better, a systematic study is needed. To accomplish this, we use each hourly gauge in the state of Texas to disaggregate all the other gauges so that trends can be identified. Three different types of gauge performance trends are investigated: (1) regional trends, where spatially-contiguous groups of gauges with equal performance are identified, (2) gauge characteristic trends, where gauges with similar statistics or model parameters are identified, and (3) proximity trends, where the distance between the disaggregated and disaggregating gauges are evaluated. In the first two cases, cluster analysis was used to identify these trends. In the third case, visual inspection was used.

Cluster analysis is an exploratory data analysis method that partitions observations into somewhat homogeneous groups within which members have similar properties compared to members in other groups (Likas et al. 2003). Most clustering algorithms fall into one of two techniques: iterative square-error partitional clustering and agglomerative hierarchical clustering (Jain et al. 2000). Among all the existing clustering algorithms, the expectation maximization (EM) algorithm–a square-error partitional clustering method–was used here because it represents the clusters in a probability-weighted fashion, so that a point has a chance of belonging to more than one cluster. The EM algorithm is a general method of finding the maximum-likelihood estimates (MLE) of parameters in probabilistic models (Bilmes 1998). The number of clusters in the model is determined by the Bayesian information criterion (BIC) (Fraley and Raftery 1998).

For each of the 323 gauges of the database and each month, the daily precipitation was disaggregated using the 323 gauges (i.e., auto-disaggregation was included), and statistics of the disaggregated time series (i.e., probability of zero rainfall, variance and lag one-hour autocorrelation coefficient) were then calculated. Thus, an array of 323 disaggregated gauges × 323 disaggregating gauges × 3 statistics × 12 months was created. Error statistics were calculated as the difference between the true statistics at a given gauge and the statistics of the time series obtained when disaggregating it using another gauge.

Regional trends

For each disaggregated gauge and month, a cluster analysis was conducted using the three error statistics of each of the 323 disaggregating gauges as cluster variables. As a result, clusters of disaggregating gauges with similar error statistics were defined. As an example, Figure 3-5 shows the clusters for disaggregating a gauge located in south Texas for the month of February. By inspection of the cluster results, the clusters did not group good-performance gauges (i.e., gauges with all error statistic values comparable to the auto-disaggregation error statistics), but rather grouped gauges of similar error statistics (e.g., high value of one error statistic and low value of another error statistic). Additionally, as evidenced in the figure, the clusters were distributed without a spatial pattern, and no regional groups of gauges with similar performance for disaggregation could be inferred from the cluster analysis.

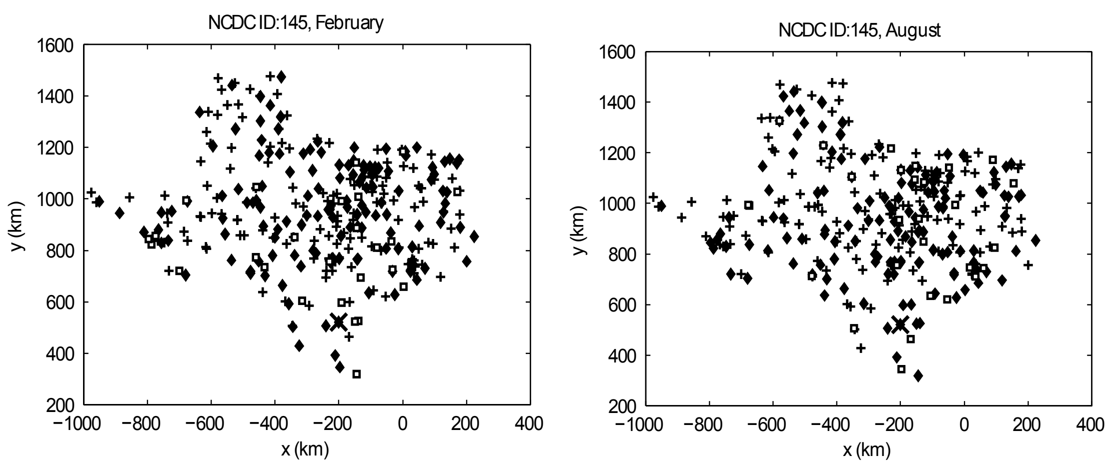


Figure 3-5. Scatterplot of clusters using the four error statistics for February (left) and August (right). No spatial or performance similarities in the clusters could be determined.

Gauge characteristic trends

The underlying assumption in the analysis of gauge characteristic trends is that good-performance gauges for disaggregating a given location have similar true statistics. Again, for each disaggregated gauge and month, a cluster analysis was conducted, but this time using the three error statistics, and the true variance and one-hour lag autocorrelation coefficient as cluster variables. The true probability of zero rainfall was not used because its range is comparatively narrower than that of the variance and the one-hour lag autocorrelation coefficient, and its use would have yielded a single cluster. As an example, Table 3-3 shows the clusters defined for disaggregating a gauge located in the Texas panhandle in August. As in the previous case, the clusters did not group together good-performance gauges, and no relationships between true statistics and error statistics was found.

Table 3-3. Cluster analysis results for evaluation of good performance gauge characteristics. The variable *m* reports the number of gauges grouped in each of the three clusters.

True σ^2	True ρ_1	Po Error	σ^2 Error	P_1 Error	m
0.00274	0.18345	-0.00424	-0.00033	0.13741	45
0.00140	0.27160	-0.00465	-0.00050	0.10578	193
0.00186	0.29784	-0.00507	0.00030	0.04948	85

Proximity trends

To evaluate whether nearby gauges perform better, the gauges with all three error statistics lower than the auto-disaggregation error statistics multiplied by 1, 1.5, 2 and 3 were plotted together with the disaggregated gauge. Figure 3-6 shows the four maps for disaggregating a gauge in south Texas in February in which the good-performance gauge databases are scattered all over Texas; no pattern with the distance to the disaggregated gauge is observed. This lack of proximity trend implies that the nearby gauge stations are not necessarily the best for rainfall disaggregation.

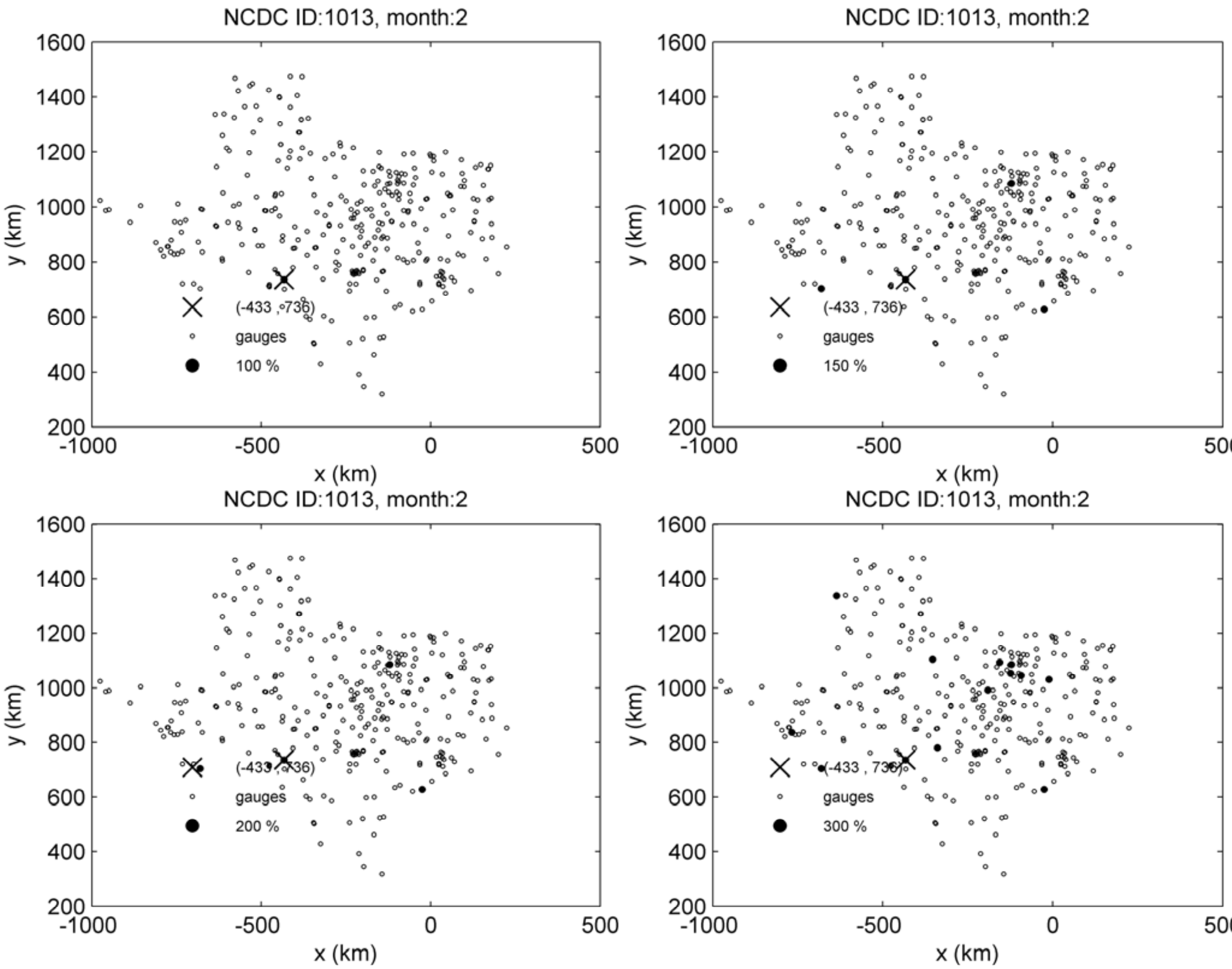


Figure 3-6. Scatterplots of databases having good performance statistics for disaggregating gauge 1013 in February.

3.3.3. State of Texas Storms Database

Based on the above trend analysis, storm databases with good disaggregation performance for disaggregating a given gauge were not located in any specific region, did not have any specific true statistic values, and were not necessarily the closest gauges. Given that there are good-performance gauges scattered throughout Texas, a Texas precipitation database that stores the precipitation data of all 323 stations was created. This database was then used to disaggregate all of the verification gauges. This database stored an overall total of 12,161 years of data and nearly 1 million storm events. The use of this Texas database implies that, regardless of the actual daily precipitation depths, when it rains, the daily precipitation variability follows common patterns for each month, and a single database can be used for disaggregation across Texas.

3.4. APPLICATION: VERIFICATION OF DISAGGREGATION METHOD

A set of 23 randomly-selected rain gauges was used for verification of the disaggregation method using the state of Texas storms database. Figure 3-7 shows an example of the performance for a gauge in southeast Texas. In the figure, the verification performance tracks well with the performance of the auto-disaggregated data. Examination of their monthly statistics shows that the model performs remarkably well on the zero-rainfall probability. Likewise, it is observed that for the variance, the verification performed better than the auto-disaggregation except for June. Both verification and auto-disaggregation had some difficulty replicating the lag one-hour autocorrelation coefficients, most of which were under-predicted. Still, the model did as well as the auto-disaggregation, implying the error is an inherited error of the method and not of the Texas database.

Figure 3-8 shows the model performance for all 23 of the verification gauges; the fit statistics are reported in Table 3-4. The performance of the state of Texas storms database and the verification is evaluated by comparison to Figure 3-4 where the results for auto-disaggregation are shown. Slightly more scatter in the probability of zero rainfall is observed in the verification, but the match is still quite close, with the r^2 values of 0.8 and 0.7 for February and August. From the slope of the regression line, the bias in the variance is significantly less in the verification, due largely to the fact that relatively low measured variance gauges were included in the verification. The autocorrelation coefficient performs similarly to the auto-disaggregation. From these verification gauges, our hypothesis that a state of Texas storms database can be used to disaggregate daily rainfall is confirmed. This fact makes application of the stochastic storm selection method straight forward, as there is no need to justify which hourly database will be used in the disaggregation. Combining the Texas storms database with the monthly values of ε from Figure 3-3, the stochastic storms selection method can be directly applied to disaggregate daily rainfall throughout Texas.

Figure 3-9 presents an example of the utility of the disaggregated data in hydrologic studies. Four representative verification gauges were selected, and intensity–duration curves are shown. For the full range of intensities in the figures, the model matches the measured values quite closely. The moderate and light intensity events below 20 mm/hr are matched nearly exactly. These are often infiltration events and are needed in water balance models to track soil moisture. For the extreme runoff-producing events, greater disagreement is observed, where the model is seen to both under-predict (gauges 587 and 4300) and over-predict (gauges 5592 and 8647) the measured values. Even so, gauges 587 and 4300 are only under-predicted by 5% and gauges 5592 and 8647 are over-predicted by 17% in the extreme events. Given that these are hourly intensity – duration curves obtained by disaggregating daily data with a one-parameter model, this agreement is promising.

Table 3-4. Performance measures between the observed and disaggregated data for the verification gauges.

Month	Statistic	MAE	*b0*	*b1*	RMSE	r^2	D
February	P^0	0.0062	0.1785	0.8183	0.0112	0.8025	0.9414
	σ^2	0.0002	0.0000	0.6837	0.0004	0.9026	0.9044
	P_0	0.1174	0.1364	0.3081	0.2044	0.2688	0.6420
August	P_0	0.0041	0.2780	0.7196	0.0075	0.8554	0.9052
	σ^2	0.0004	0.0004	0.6536	0.0007	0.8183	0.9161
	ρ_1	0.1053	0.0916	0.2297	0.1621	0.0477	0.4466

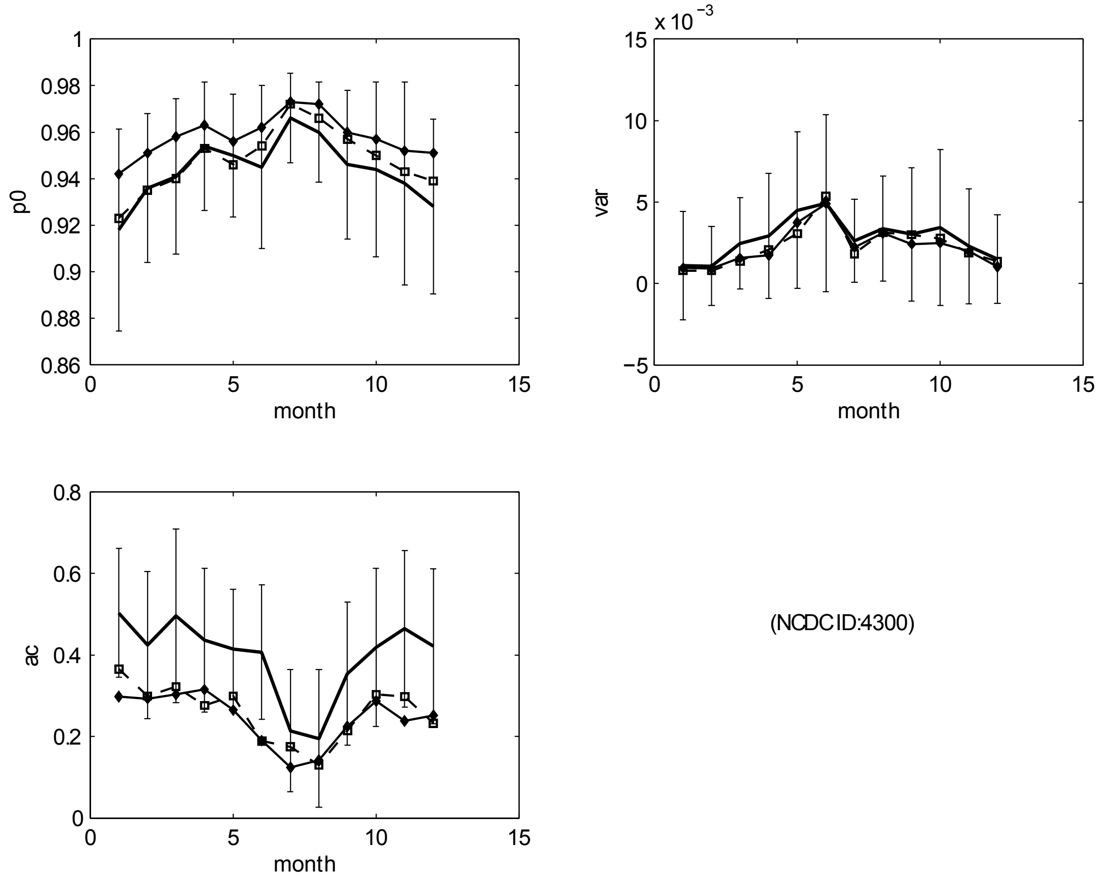


Figure 3-7. Performance of the disaggregation model for gauge 311. Legend – The solid line is the measured statistics, the line with open boxes is the auto-disaggregation model statistics, and the line with solid diamonds is the disaggregation model statistics for the state of Texas storms database. The error bars show the monthly variability in the measured data.

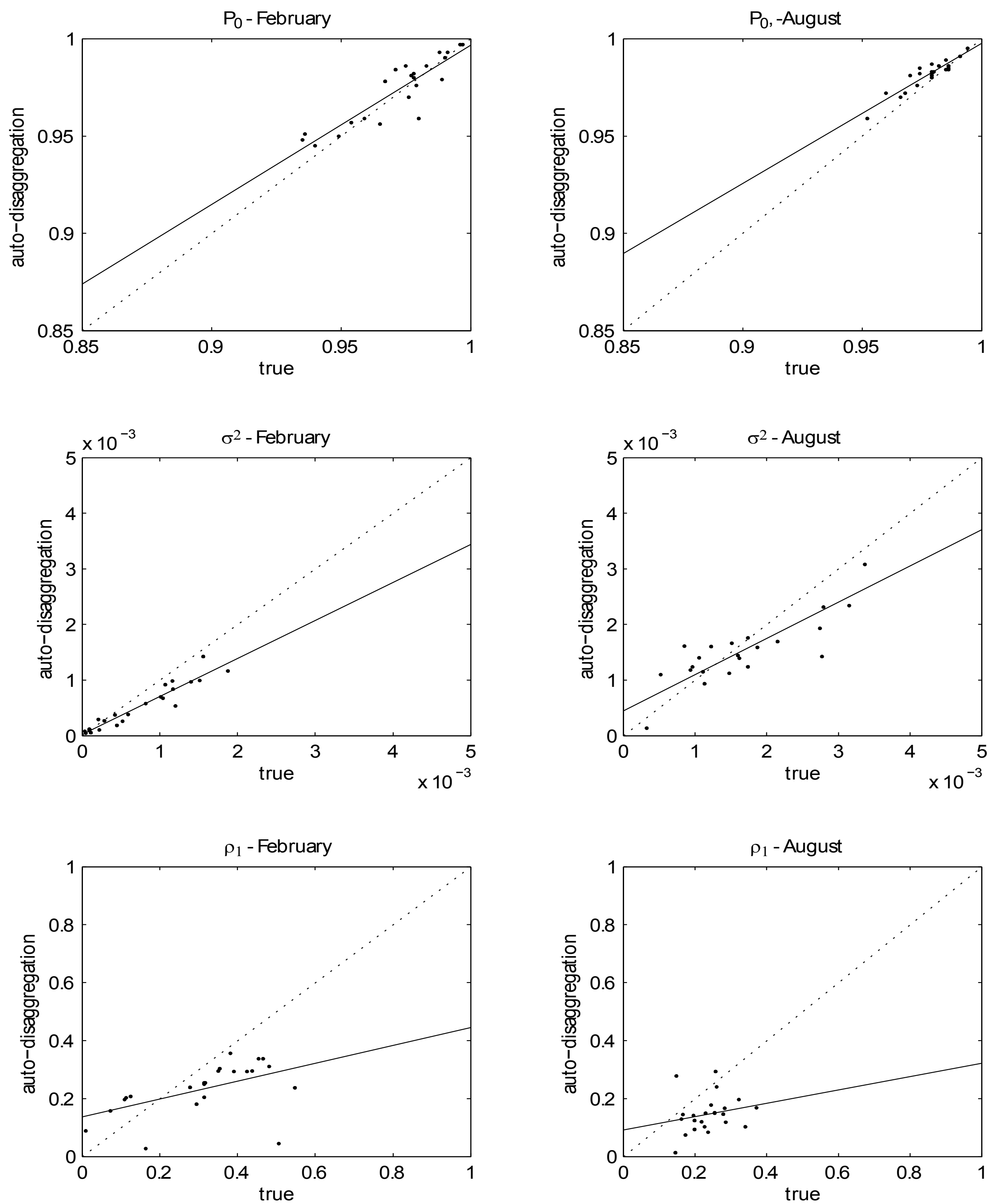


Figure 3-8. Comparison between general disaggregation and true statistics using the state of Texas storms database. The solid line represents a linear regression model. The dashed line represents the 1:1 line (a perfect model).

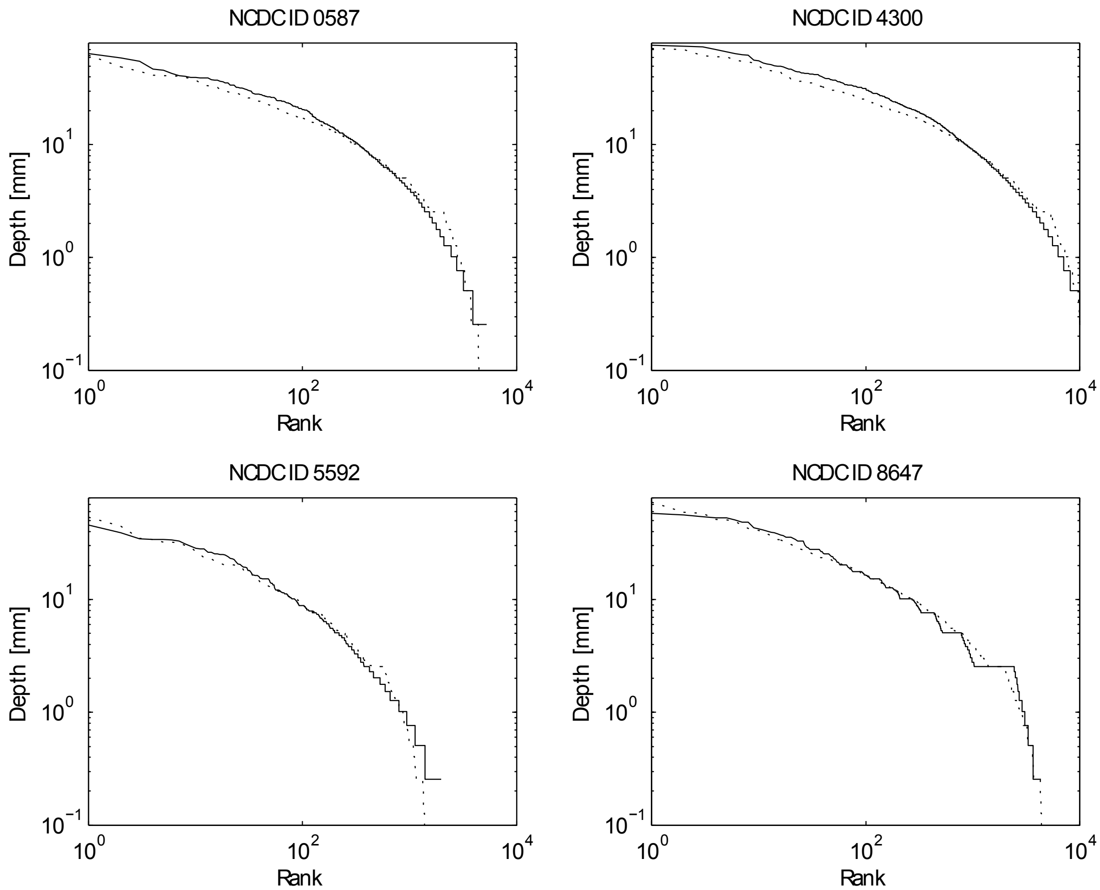


Figure 3-9. Intensity duration curves for four of the validation gauges using the state of Texas storms database.

3.5. SUMMARY AND CONCLUSIONS

The stochastic storm selection method has been evaluated for disaggregating daily data into hourly precipitation across Texas. The auto-disaggregation of historic hourly rainfall in Texas provides both the calibrated value of the smallest hourly event ε in Texas and a guideline for the best expected performance of the model. By spatial fitting of the ε parameter values, no significant spatial trends are found. Moreover, ε follows a seasonal variability from low values in the winter corresponding to frontal system storms and higher values in the summer corresponding to convective thunder showers. The auto-disaggregation performance is good for the probability of zero rainfall and for the variance for gauges with low measured values of variance. Both high variance and high lag one-hour autocorrelation coefficient are systematically underestimated by the method.

Selection of the appropriate hourly database to use in disaggregation was evaluated using cluster and spatial data analysis. Using the auto-disaggregated data, no trends were found either in the spatial distribution of good databases or in their modeled or measured characteristics (e.g. ε values or rainfall statistics). Based on this analysis, we conclude that a single hourly database containing all the measured precipitation in Texas can be used for disaggregation. Several gauges were selected for verification, and these gauges confirmed that the state database performed as well as the auto-disaggregation.

From a practical point of view, this study provides the analysis needed to apply the stochastic storm selection method across Texas and the guidance needed to evaluate other locations. The average monthly values of ε can be used for the single model parameter irrespective of location; the state of Texas storms database can be used as the hourly off-site database. When applied in this way, the performance of the model can be expected to lie in the range of the auto-disaggregation. When this method is applied across the United States, we expect to find a limited number of similar regional databases and similar large regions of uniform ε values.

When applied to a hydrologic study, the disaggregation scheme provides very good intensity–duration curves. The curves match very closely the majority of moderate rainfall events. This match is critical for accurate tracking of soil moisture in water balance models. The curves both over- and under-predict the extreme runoff-generating events, but they are still matched to within 17% of the measured values.

4. STORM IDENTIFICATION AND TRACKING ALGORITHM FOR MODELING OF RAINFALL FIELDS USING 1-HOUR NEXRAD RAINFALL DATA IN TEXAS

A method to identify and track rainfall structures using one-hour accumulated NEXRAD rainfall data is presented and used to analyze the dynamics of storm features over an area in Texas. Storm features are identified from a Gaussian mixture model using the expectation maximization algorithm. The method assigns NEXRAD pixels to storm features while simultaneously producing a smooth fitted function to the rainfall intensity distribution. Once the storm features are identified, they are tracked using inverse cost functions and using the fact that continuous features overlap each other from frame to frame in the accumulated data. The inverse cost functions also account for storm feature merging, splitting, birth, and death. Application of this storm identification and tracking algorithm for Brazos County (1,500 km^2) in southeastern Texas distinguishes several characteristics of the storm feature dynamics. From September through April, storm features are predominantly frontal in nature, with storm features following geostrophic flow along low pressure fronts moving in from the north. In summer (May-August), storm features are convective in nature following random track directions. Both types of storm features have durations of one to three hours in Brazos County due to the county's relatively small size compared to the measured average storm speed of 40 km/hr and due to the fact that most storms only intersect the county over part of their area.

4.1. INTRODUCTION

Watershed responses to storm events are strongly affected by the spatial and temporal patterns of rainfall; that is, the spatial distribution of the precipitation intensity and its evolution over time. Although real storms move and have non-uniform intensity distributions in both space and time, hydrologic applications often simplify their dynamics by assuming storms that are uniformly distributed and have variable intensity in time according to a pre-defined hyetograph shape. As one considers watersheds of greater size, the non-uniformity of rainfall becomes increasingly important because a storm may not cover the watershed's entire area (Olivera et al. 2006) and may not stay in the watershed for its full duration. In order to incorporate parameters such as area, shape, velocity, and distributed intensity field in the definition of synthetic storms, it is necessary to determine these characteristics from spatially distributed precipitation data. To date, most algorithms for identifying and tracking storm characteristics have been applied to short time-step radar reflectivity data (e.g., 15 minutes or less), where storm features are captured in an effectively synoptic manner. Throughout United States, however, the most reliable data for distributed precipitation are the one-hour accumulated next generation radar (NEXRAD) data of the U.S. National Weather Service (Wilson and Brandes 1979; Crum and Alberty 1993; Crum et al. 1993; Klazura and Imy 1993; Brown and Lewis 2005; National Weather Service 2006). The one-hour aggregation level of the data makes it more difficult to identify and track storms than when using sequences of synoptic radar reflectivity data because storms can traverse over a number of NEXRAD pixels and can change size and shape appreciably between consecutive data maps. In this dissertation, we present a methodology to overcome these identification and tracking difficulties and to extract the characteristics of moving storms from one-hour accumulated distributed rain-fall data. This method is important to determine the non-uniform characteristics of storms over a region of interest such as a watershed, city, county, state, or other geographic feature of the United States using the publicly available NEXRAD precipitation data.

A description of the hierarchical structure of rainfall fields was introduced by Le Cam (1961), who also proposed the utility of using cluster-point processes to explain the spatial and temporal characteristics of storms. In general, distributed rainfall models have utilized this description of storm structure to develop models along three basic lines (Rodriguez-Iturbe et al. 1998): multifractal precipitation models based on random cascades (e.g., Lovejoy and Schertzer 1986; Gupta and Waymire 1993; Tessier et al. 1993; Seed et al. 1999; Lovejoy and Schertzer 2006), spatial random function generators (e.g., Bellin and Rubin 1996), and stochastic cluster-point process models (e.g., Rodriguez-Iturbe et al. 1986; Cox and Isham 1988; Cowpertwait 1995; Northrop 1998; Wheater et al. 2000). These methods have also been developed for simulating or disaggregating rainfall at a point (e.g., Woolhiser and Osborn 1985; Hershenhorn and Woolhiser 1987; Rodriguez-Iturbe et al. 1987; Rodriguez-Iturbe et al. 1988; Econopouly et al. 1990; Bo et al. 1994; Cowpertwait 1994; Veneziano et al. 1996; Cowpertwait 1998; Olsson 1998; Socolofsky et al. 2001; Choi et al. 2006). As a related approach, Krajewski et al. (1993) use a stochastic approach to simulate radar reflectivity and then convert these synthetic radar images to rainfall fields.

To analyze the performance of rainfall models and to evaluate their parameters, analytical and/or numerical properties of measured and simulated distributed precipitation are derived and compared. These properties include statistics at a point, at multiple points, and of distributed measurements. For instance, Cowpertwait (1995) compares multi-site, second-order properties between measured and predicted rainfall for a cluster-point process model with circular rain cells. Northrop (1998) likewise extends Cowpertwait's (1995) analysis to include elliptical rain cells. From the perspective of starting with the measured data, Rodriguez-Iturbe et al. (1998) derive metrics for evaluating synthetic space-time rainfall generators. Using daily accumulated radar data, they show that the exponent of the scaling relationship for spatial variance over time has a characteristic value that should be matched by simulations. A few studies use short-term weather data to determine the stochastic attributes of storms. One such study is reported by De Lannoy et al. (2005), who use short-time radar data to characterize the size and structure of storms. They show that structure scaling laws are fit better by using a local coordinate system oriented in the instantaneous direction of the storm movement. Overall, all of these rainfall simulation approaches benefit from detailed knowledge of the distributed characteristics of storm structures.

For distributed rainfall data to be used to develop physically based models, a means is needed to classify the rainfall structures objectively. Veneziano and Villani (1996) apply statistical models to radar data to identify the storm structures. They use the expectation maximization algorithm with Gaussian mixture models to assign radar pixels to storm cells by a probabilistic approach. The method allows rain cells, which are clusters of related reflectivity, to be rigorously identified. While Veneziano and Villani (1996) did not track these clusters over time, each radar image is independently analyzed to obtain spatial statistics of the rainfall process at fixed time intervals. Other authors also study the spatial structure of rainfall in radar data (Rodriguez-Iturbe et al. 1998; Schumacher and Johnson 2005; Schumacher and Johnson 2006). While important for understanding the rainfall structures, these identification procedures are applied for the purpose of classifying pixels into storm cells and do not result in fitted models that represent the measured rainfall intensity field.

In addition to rainfall modeling, another objective of analyzing spatial rainfall data is to improve short-term weather forecasting, or nowcasting. Many tracking and extrapolation algorithms have been developed for this purpose. The tracking algorithms are applied to radar reflectivity data and can be classified into three main methods: spatial autocorrelation (e.g., Zawadzki 1973; Rinehart and Garvey 1978; Crane 1979; Li et al. 1995; Mecklenburg et al. 2000b; Upton 2000; Dell'Acqua and Gamba 2002), direct tracking of identified rain cells (e.g., Dixon and Wiener 1993), and hybrid methods that take advantage of both autocorrelation and direct tracking (e.g., Einfalt et al. 1990; Bremaud and Pointin 1993; Johnson et al. 1998; Grecu and Krajewski

2000; Steinacker et al. 2000; Handwerker 2002; Lakshmanan et al. 2003; Bowler et al. 2004). Auto-correlation has been applied for over 40 years (Wilson et al. 1998), and can be applied to a full-field scan to get an average storm velocity (e.g., Zawadzki 1973) or to interrogation sub-windows to get full-field velocity information (e.g., Rinehart and Garvey 1978), similarly to the particle image velocimetry method applied in experimental fluid mechanics (Adrian 1991; Raffel et al. 1998). The difficulty of applying autocorrelation to radar data is the fact that storms change size, shape, and intensity on a time scale shorter than that at which synoptic scans are obtained. The direct tracking algorithms overcome some of this difficulty by first identifying structures to track (e.g., rain cells) and then linking structures from scan to scan using cost functions. The cost function can be as simple as taking the nearest neighbor approach or more complex, as in hybrid approaches that consider the flow field implied by the autocorrelation method and that account for storm merging, splitting, growth, decay, birth, and death. Forecasts are made by extrapolating various levels of physically-based models conditioned on the known storm cell tracks. While these methods can reliably track storm structures to produce accurate short-term forecasts, they have not been applied to track structures in one-hour accumulated precipitation fields.

An important hydrologic application of storm tracking is to use nowcasting of spatial precipitation data to improve real-time forecasting. The rainfall data are used as input to distributed flood forecasting models, and comparisons are made between simulations using uniform and distributed data and between simulations using the measured and forecasted precipitation (e.g., Mecklenburg et al. 2000a; Berenguer et al. 2005; Morin et al. 2006). These simulations confirm that spatially variable rainfall can produce quite different results from uniformly distributed rain-fall in distributed flood models. In addition, peak flood flows are sensitive to the location and movement of rainfall over an area. For instance, Morin et al. (2006) showed that a storm moving parallel to the main channel flow in Walnut Gulch Experimental Watershed (148 km^2 in southern Arizona) would double peak flows if the storm track had originated 2 km closer to the river than was recorded. In all cases, simulations using forecasted distributed rainfall produced better real-time flood predictions than those based on uniform precipitation. However, these simulations are sensitive to the locations of rainfall, and it remains to be learned how to include distributed, dynamic synthetic storms in hydrologic applications.

In order to obtain basic information on the spatial and dynamic patterns of rainfall over an area, we present a storm identification and tracking algorithm that uses the one-hour accumulated NEXRAD precipitation data of the U.S. National Weather Service. The method uses a similar algorithm to Veneziano and Villani (1996) to identify storm features (i.e., defined structures within a storm), but uses a Gaussian mixture model to also obtain an actual fitted curve representing the distributed rainfall intensity. The tracking algorithm overcomes the difficulties of tracking one-hour accumulated rainfall by observing that storm features along continuous tracks overlap from image to image. An inverse cost function is derived to identify the tracks of storm features and to address merging, splitting, birth, and death. The Methodology section presents the details of the storm identification and tracking algorithm. The Application, Results, and Discussion section presents an analysis of the results for storms traversing Brazos County in south-eastern Texas in 2003.

4.1.1. NEXRAD Data

In the United States, the National Weather Service (NWS) distributes radar-based precipitation depths that are estimated as part of the NEXRAD federal program. This program has deployed a number of S-band weather surveillance 1988 Doppler (WSR-88D) radars across the country. WSR-88D raw data are collected as radar reflectivity on a polar grid. These data are subsequently converted to precipitation estimates on a rectilinear grid. The precipitation estimates based on multiple data sources for calibration and verification (e.g. radar, rain gauges, and satellite imagery) are publicly distributed as multisensor precipitation estimator (MPE) data. The MPE rain estimates are available as digital precipitation arrays (DPA) mapped onto the national hydrologic rainfall analysis project (HRAP) grid. The HRAP grid pixels are square with a side length of 4 km, but when projected to a specific location, they remain nominally square with side lengths varying from about 3.7 km at southern U.S. latitudes to about 4.4 km at northern U.S. latitudes (Fulton et al. 1998). Detailed discussion of the development of radar-based precipitation depth maps, sources of error and algorithms used for adjusting precipitation estimates to observed data is presented in Fulton et al. (1998), Brown and Lewis (2005), Crum and Alberty (1993), Crum et al. (1993), Klazura and Imy (1993), Wilson and Brandes (1979), Seo et al. (1999), Nelson et al. (2006), Kondragunta et al. (2005), Reed and Maidment (1999), among others. Notwithstanding the fact that NEXRAD precipitation data are subject to inaccuracies due to a number of error sources, it should be acknowledged that, at present, it is the best continuous distributed precipitation database available in the United States at its resolution.

4.2. METHODOLOGY

The methods developed here for storm identification and tracking overcome difficulties caused by the averaging over one-hour accumulation periods and over 16 km^2 areas of the NEXRAD precipitation data. Nevertheless, these data contain meaningful information on the storm features. In this study, storm features are defined as clustered areas of high precipitation, but due to the accumulation of the data over an hour, they do not directly relate to hierarchical storm structures, such as rain cells. Storm features are identified in the one-hour accumulated rainfall images using thresholds on rainfall depth and feature size and Gaussian mixture models, which are also used to fit a smooth surface to the measured rainfall intensity. Once storm features are identified and fitted, they are matched and related to their counterparts in the ensuing time intervals to determine their dynamic characteristics.

4.2.1. Storm Identification and Fitting

The procedure for identifying storm features consists of applying a low-pass filter to the rainfall image, segmenting the rainfall image into precipitation zones using contour lines of given intensity value, and segmenting the zones into storm features using Gaussian mixture models (GMM). GMMs are also used for fitting the distribution of precipitation intensity in the storm features.

To smooth the rainfall image and remove unusually high or low intensity values, a median low-pass filter is used. The filter replaces the intensity values at each pixel with the median value in its vicinity, which is a moving window of predefined shape and centered at the pixel. After smoothing, the filtered image captures the general precipitation patterns and not the data particulars. For illustration purposes, the raw and filtered images of Brazos County, Texas, for October 5th, 2003, 10:00 pm – 11:00 pm, processed with a 3×3 pixel vicinity, are shown in Figure 4-1. The window shown in Figure 4-1 corresponds to the study area discussed in the Application, Results and Discussion section.

After filtering the image, it is segmented into zones using contour lines of a given intensity value δ (refer to Figure 4-2). The areas of these zones are then compared to a predefined range of acceptable storm feature sizes. Those zones smaller than the minimum size A_{min} are removed, and those larger than the maximum size A_{max} are further segmented with contour lines of intensity $\delta + \Delta\delta$. This process is repeated with intensities of $\delta + 2\Delta\delta$, $\delta + 3\Delta\delta$ and so on, until all the identified zones fall within the predefined range of acceptable storm feature sizes. When determining this range, the data resolution impacts A_{min}, and the computational efficiency impacts A_{max}.

After segmentation by contouring, the resulting zones are further processed to identify storm features and fit their intensity patterns using GMMs. GMMs have been used in the past in different disciplines to find patterns of interest in random fields (e.g., Redner and Walker 1984; Blimes 1998; Fraley and Raftery 1998). Veneziano and Villani (1996), in particular, use GMMs to cluster irregular-shaped regions of radar reflectivity to identify storm cells, and represented them as linear combinations of Gaussian distributions. In their linear combination of distributions, the probability density indicates the probability that a given pixel belongs in a cluster. In our model, GMMs are used not only for identifying storm features (i.e., assigning pixels to clusters) in the already segmented zones, but also for representing the pixel precipitation intensity using linear combinations of Gaussian distributions.

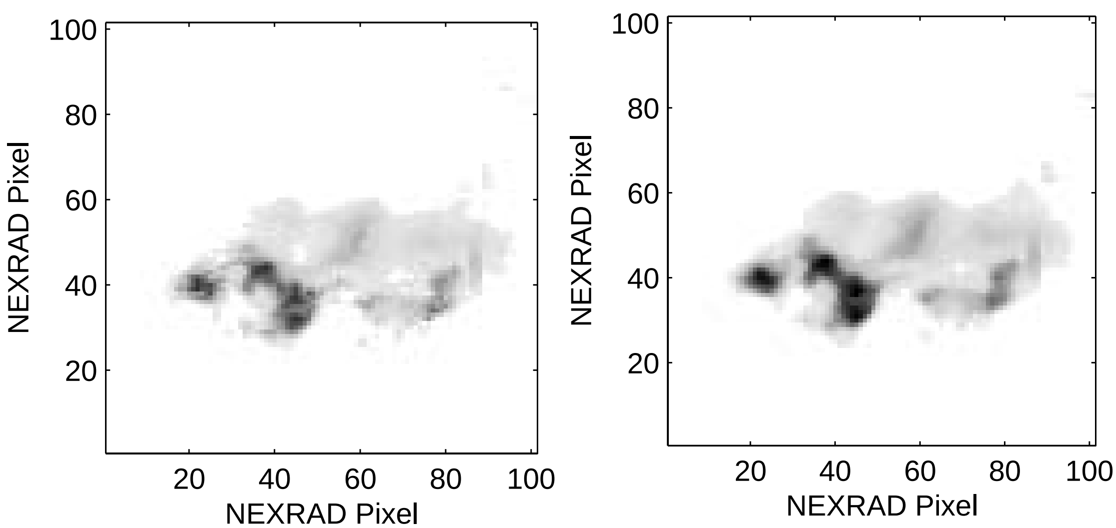


Figure 4-1. Original image (left) and filtered image (right) using a median filter.

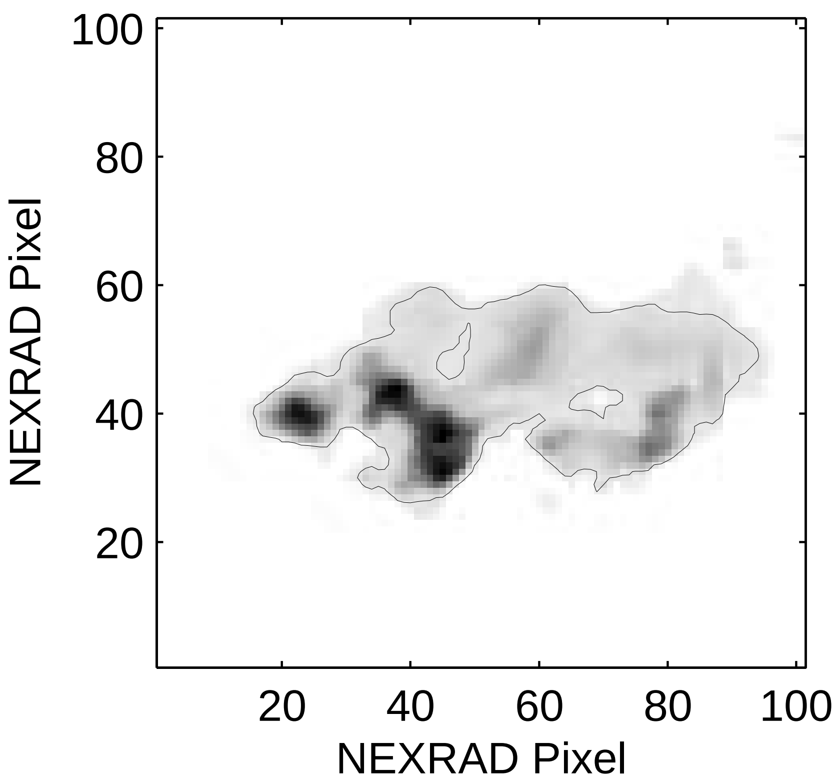


Figure 4-2. Image segmentation using contouring.

Each segmented zone is assumed to be represented by a mixture of N Gaussian bivariate distributions, or components, as follows

$$P(x \mid \Theta) = \sum_{i=1}^{N} \omega_i \, p(x \mid \theta_i) \tag{4-1}$$

where $P(x \mid \Theta)$ is the density function of the Gaussian mixture, $p(x \mid \theta_i)$ is the density function of component *i*, x is the (2×1) location vector, and $\Theta = \{\omega_1, ..., \omega_N, \theta_1, ..., \theta_N\}$ contains the distribution parameters. Specifically, ω_i is the mixing weight of component *i*; and $\theta_i = \{\mu_i, \Sigma_i\}$ contains the mean vector μ_i and the covariance matrix Σ_i. The mixing weights satisfy the condition $\sum_{i=1}^{N} \omega_i = 1$. The individual components are expressed as:

$$p(x \mid \theta_i) = \frac{1}{2\pi \Sigma_i^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right] \tag{4-2}$$

To define the Gaussian mixture of a given zone, the parameters *N*, ω_i, μ_i and Σ_i must be determined. The expectation maximization (EM) algorithm (Redner and Walker 1984; Blimes 1998) is used to determine the maximum likelihood parameters of the Gaussian mixture for a given number of components. The number of components is determined by the Bayesian information criterion (BIC) (Fraley and Raftery 1998), which indicates

$$BIC(N) = 2L(\hat{\Theta}) - m_N \ln(n) \tag{4-3}$$

where $\hat{\Theta}$ are the estimates of the parameters, $L(\hat{\Theta})$ is the logarithm of the likelihood of the estimated parameters, m_N is the number of model parameters, and *n* is the sample size. The value of N is determined iteratively. First, a value *N* = 1 is assumed, and the parameters $\hat{\Theta}$ and likelihood $L(\hat{\Theta})$ are calculated with the EM algorithm, and BIC(*N*) is determined. Next, *N* is increased by one, and again the parameters $\hat{\Theta}$, the log-likelihood $L(\hat{\Theta})$ and BIC(*N*) are determined. The optimum value of *N* is selected to maximize BIC(*N*).

Unlike previous applications of GMMs which assign pixels to clusters, we desire to also fit the GMM to the precipitation depth distribution. To accomplish this, the rainfall depth at a pixel in a zone is represented as an integer number of elementary depth blocks.

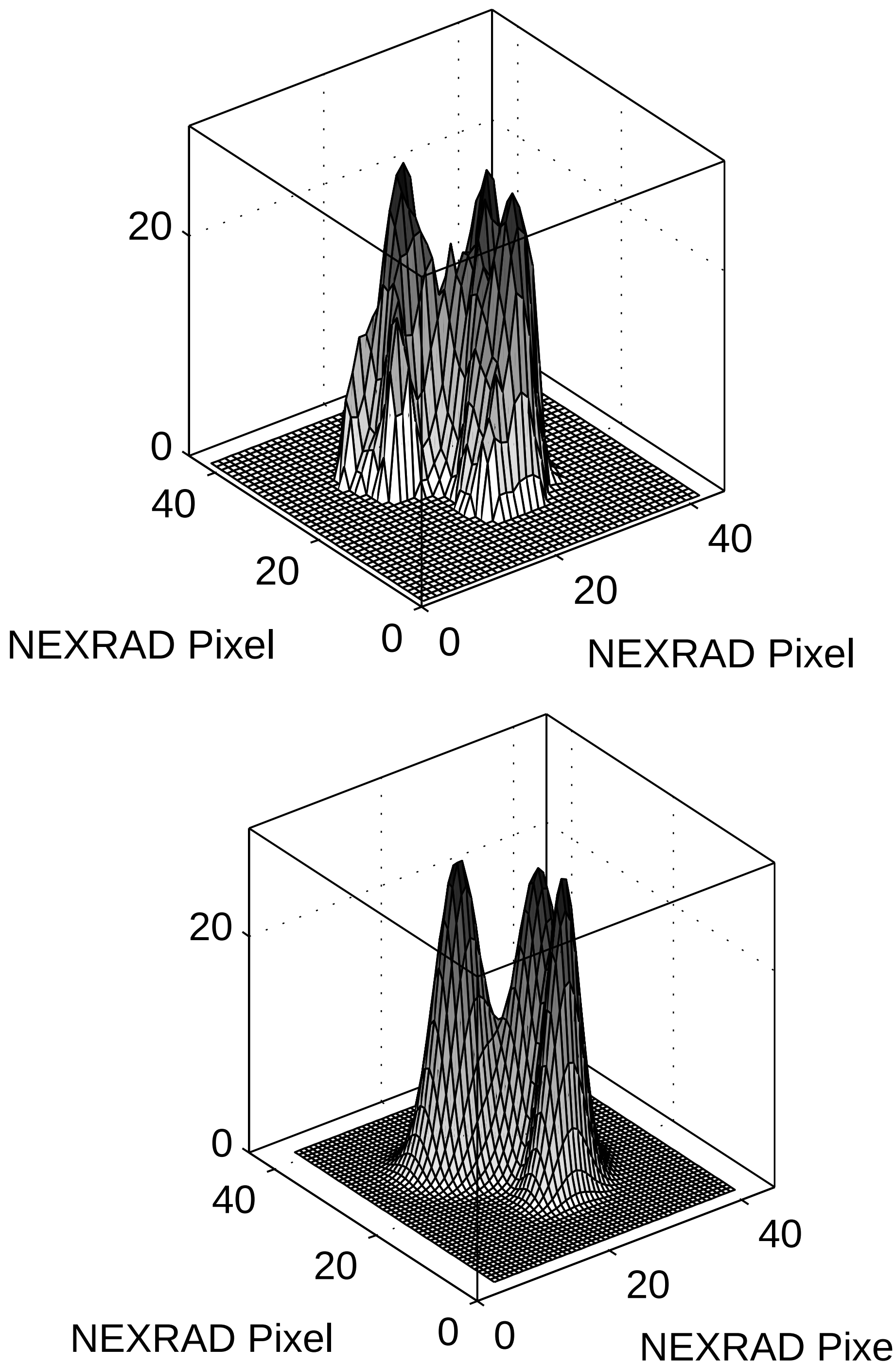
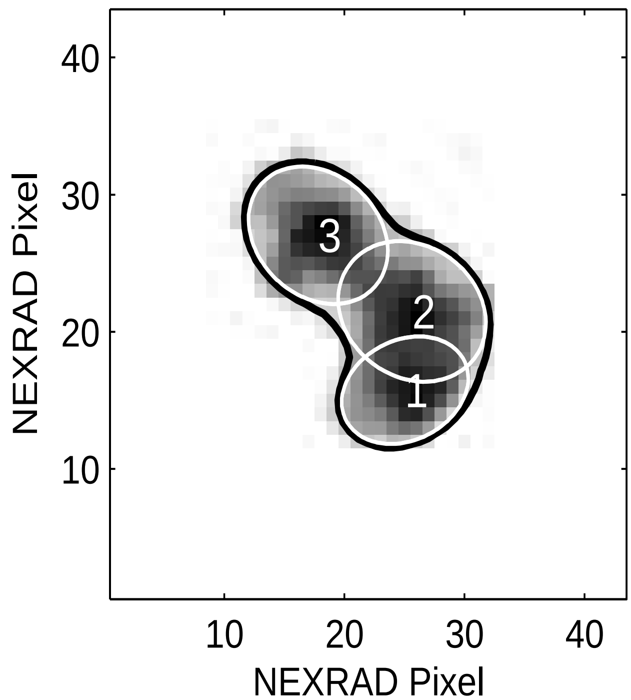


Figure 4-3. Precipitation depth in mm of a storm feature: original image (top) and GMM fitted image (bottom).

The number of blocks k at a pixel is proportional to the observed rainfall depth D such that κ=int(D/κ) where κ is the elementary rainfall depth. Each pixel is then represented k times in the GMM so that the sample size n in the EM algorithm is equal to the number of elementary depth blocks in the zone. The smoothed intensity distribution and the fitted GMM for a single feature of the storm in Figure 4-1 is shown in Figure 4-3.

Once the components are defined, the zones are further segmented into storm features based on the overlapping areas of the components (refer to Figure 4-4). To make the component areas finite, they are truncated at a given quantile q. After this truncation, the components either intersect each other or are isolated. By setting a non-dimensional overlapping area threshold A_0, storm features are taken as an individual or as a combination of components. That is, if the overlapping area divided by the area of the smaller component is less than A_0, then the components are taken as separate storm features; otherwise, the components are taken as a single storm feature. For the zone in Figure 4-5, all three components were assigned to a single storm feature. Figure 4-5 shows the eight storm features into which the zone in Figure 4-2 is segmented. The overlapping features to the east in Figure 4-5 are separate features because their overlapping areas are smaller than the pre-defined threshold. Thus, the storm feature identification and fitting requires seven parameters plus the definition of the vicinity size and shape for the low-pass median filter.



Storm feature with 3 components

Component 1: (ω=24.9 %)

$\boldsymbol{\mu}$ = (25.1, 15.7)

$$\boldsymbol{\Sigma} = \begin{bmatrix} 6.0 & 0.996 \\ 0.996 & 3.33 \end{bmatrix}$$

Component 2: (ω=38.5 %)

$\boldsymbol{\mu}$ = (25.6, 21.5)

$$\boldsymbol{\Sigma} = \begin{bmatrix} 8.17 & -1.14 \\ -1.14 & 5.68 \end{bmatrix}$$

Component 3: (ω=36.6 %)

$\boldsymbol{\mu}$ = (17, 8.27) $\boldsymbol{\Sigma} = \begin{bmatrix} 7.27 & -1.43 \\ -1.43 & 5.44 \end{bmatrix}$

Figure 4-4. GMM of a single storm feature.

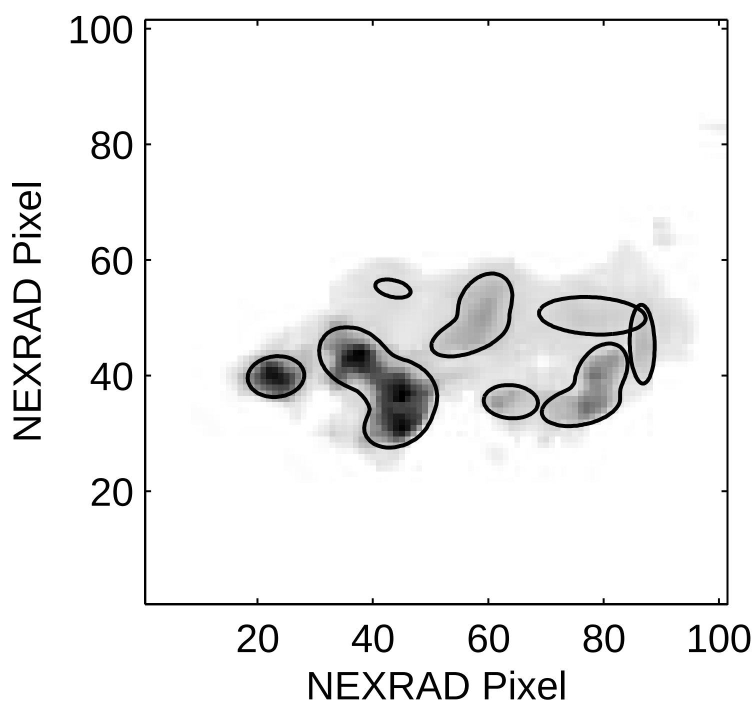


Figure 4-5. Storm feature segmentation using GMM.

4.2.2. Storm Tracking

Since the NEXRAD data stores precipitation depth accumulated over a one-hour interval (i.e., it is not a synoptic snapshot of rainfall intensities), moving storm features are assumed to overlap on consecutive time frames. Therefore, for tracking a given storm feature, those features in the ensuing time frame that overlap the original feature are identified and their degree of association with it is evaluated based on inverse cost functions. In general, these inverse cost functions utilize geometric properties of the storm features and of their intersections as well as prior knowledge of the storm track velocity field.

The storm feature and intersection properties are evaluated by comparing two consecutive images, called an image pair. For single images, each storm feature is characterized by its area and centroid. For the image pairs, the intersecting regions of storm features between two images are found; for each intersection, the overlapping area between two storm features is calculated.

Using these geometric properties, the degree of association between overlapping storm features is evaluated using an inverse cost function. For the intersection of storm features presented in Figure 4-6, the non-intersecting area at the initial time is A, the intersected area is B, and the non-intersecting area at the later time frame is C. Then, the inverse cost function is given by

$$S = f\phi + (1 - f)\psi \tag{4-4}$$

where ϕ is the dimensionless intersected area, given by

$$\phi = \frac{1}{2}\left(\frac{B}{A+B} + \frac{B}{C+B}\right) \tag{4-5}$$

ψ is the dimensionless difference between intersecting storm areas, given by

$$\psi = 1 - \frac{|(A+B)-(C+B)|}{\max[(A+B),(C+B)]} \tag{4-6}$$

and f is weighting factor, taken as a model parameter. The storm features at consecutive times are assumed to be the same feature if S is above a threshold value S_{min}.

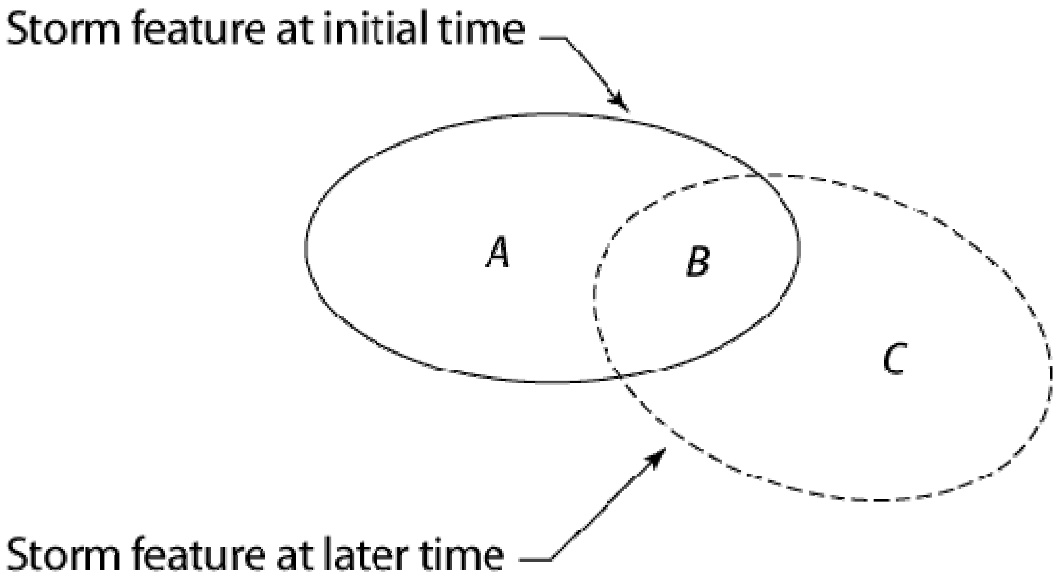


Figure 4-6. Diagram of a storm feature at two time showing the non-intersected areas *A* and *C* and their intersected area *B*.

The inverse cost function defined in equation (4-4) specifies the degree of agreement between storm features in consecutive time frames. If the intersecting area B much larger than A and C, the storm is moving slowly with respect to its size. In this case, ϕ and ψ approach one and S approaches its maximum value of one. On the other hand, if the intersecting area B is much smaller than A and C, and A and C are similar in size, the storm feature may be a single feature that moves quickly relative to its size. In this case, ϕ approaches zero while ψ approaches one so that S approaches (1-f). Finally, if the intersecting area B is small compared to A and C and either A or C is large compared to the other, the feature is either growing or decaying, or the two storm features are not related. Here, ϕ, ψ, and S approach zero. Thus, the threshold value S_{min} together with the weighting factor f specify the criteria to evaluate whether a storm feature is moving too quickly, changing its size (growth or decay) or whether a new feature is appearing.

At the beginning of each rainy period, no prior storm track velocities are known, and the tracking procedure is initiated using the cost function in equation (4-4) only. Initially, storm tracks are identified only for features that intersect a single feature at the subsequent time. Once these feature pairs are identified, the inverse cost function is evaluated to determine whether these features should be considered a continuation of a previous storm feature. The storm features are considered linked if the inverse cost function evaluates above a threshold S_{min1} using a weighting factor f_1. If all initial features have multiple pairs, then the linked pairs having a maximum value of the inverse cost function above the threshold value are assumed to be the correct pairs.

Once the continuing, or linked, storm features are identified, a velocity can be calculated by vector arithmetic. The centroid of each storm feature at each time interval is calculated as the weighted average of the Gaussian mixture model for that storm feature. The velocity vector components are then the displacement of the centroids of the storm features in the Cartesian directions x and y divided by the time interval (one hour).

To track the remaining storm features at the initial time, the known velocity vectors are used. First, the known velocity field is interpolated to the centroids of the storm features using a linear (inverse distance) interpolation to obtain velocity estimates $\hat{V}$. Velocity vectors are interpolated only from known vectors within a circle of a specified radius R. Second, a search window for possible continuing storm features is defined by a maximum deviation from the interpolated velocity direction ε and a maximum displacement L. The maximum displacement is estimated from an absolute maximum $L_{\max} = V_{\max} \Delta t$ or a local estimate of the displacement

$L_{local} = \lambda \, | \hat{V} | \, \Delta t$, where V_{max} is the maximum expected storm feature velocity and λ is a buffer factor to account for the fact that the velocity could be greater than the interpolated value. Here, we take L to be the minimum of these two estimates. Intersected storm features whose centroids lie outside this region are either born at the subsequent time or possibly associated with other storm features in the initial time frame. Those intersected storm features inside this region are designated as possible storm feature tracks. Third, the likelihood that features continue along a storm track is evaluated by the inverse cost function in equation (4-4) with new threshold values S_{min2} and weight f_2. In principle, these thresholds may be different than those specified above because we can make use of the information in the known velocity field and can specify stricter criteria for associating storm features. Storm features split when multiple tracks diverging from a single initial storm feature have S above S_{min2}; storm features merge when multiple initial storm features track to a single subsequent storm feature. Storm features are born or die when no tracks are identified according to the cost function.

Within rainy periods, velocity field information is known at previous time steps. Thus, the algorithm uses the search window defined by the interpolated velocity vectors to evaluate all storm tracks. The inverse cost function is evaluated using the weight f_2 and compared to the threshold value $S_{\min 2}$. This procedure continues until an hour of dry weather in encountered. The algorithm is re-started at the beginning of a new rainy period, as discussed above, ignoring any previous velocity fields. Thus, the storm feature tracking algorithm uses seven parameters

4.3. APPLICATION, RESULTS, AND DISCUSSION

As an illustration of this storm identification and tracking algorithm, storm features from NEXRAD data for a portion of Texas for 2003 are analyzed. The purpose of this application is to find the patterns of storm dynamic characteristics over Brazos County.

4.3.1. Data and Study Area

The precipitation data used are the West Gulf River Forecasting Center (WGRFC) NEXRAD Multisensor Precipitation Estimator (MPE) data for calendar year 2003 (National Weather Service 2006). These data had six missing days and a few isolated missing hours leaving a total of 8,613 hours of recorded data. The data were clipped by a rectangular window consisting of 101×101 NEXRAD cells (approximately 151,000 km^2) centered on Brazos County, Texas (refer to Figure 4-7). Brazos County has an area of 1,500 km^2. It is bounded to the south-west by the Brazos River, to the east by the Navasota River, and to the north-west by Old San Antonio Road (OSR).

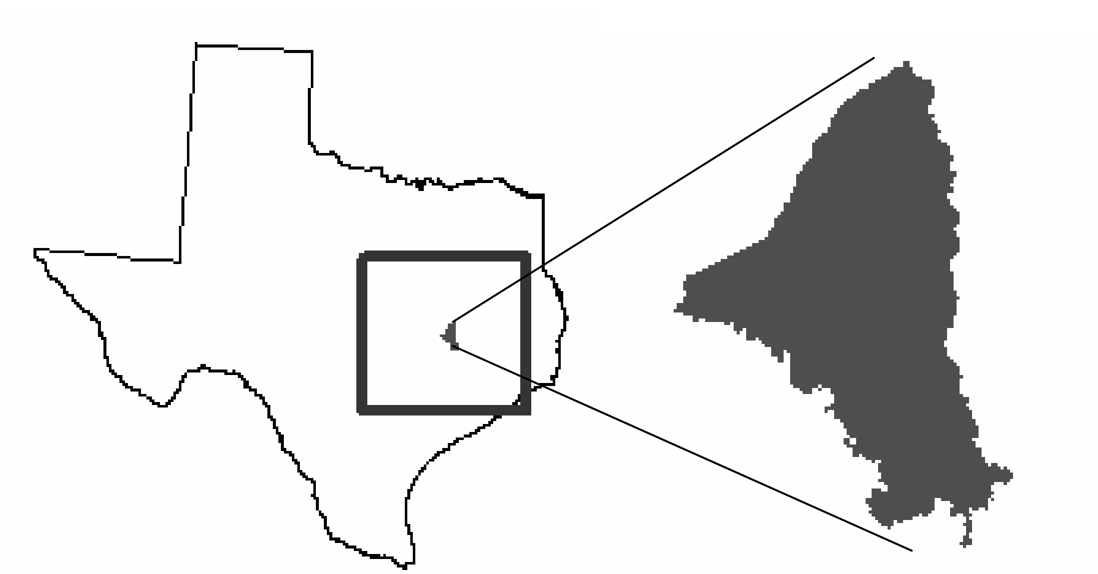


Figure 4-7. Brazos County in southeast Texas and the 101 × 101 NEXRAD pixel window (left) and an enlarged view of Brazos County (right).

4.3.2. Model Parameters

The model parameters for both the storm feature identification and tracking algorithms are presented in Table 4-1. The maximum storm feature speed is taken from recommended values (Dixon and Wiener 1993); all other parameters are specific to our method and were determined by sensitivity analysis and visual optimization. For storm feature identification, the contour level δ and incremental contour interval $\Delta\delta$ determine the number of individual storm features and are the most sensitive parameters. For storm tracking, the weighting factor f and the inverse cost function threshold $S_{\min}$ work together to determine storm feature birth, death, splitting, and merging. This parameter set is deemed to be robust because a different set of parameters changes the resolution of the tracked features but does not result in false storm tracks.

Table 4-1. Parameter set used for the identification and tracking algorithm for the application to the area around Brazos County.

Parameter	Value
δ	1 mm
$\Delta\delta$	0.5 mm
$\begin{bmatrix} A_{\min} & A_{\max} \end{bmatrix}$	$\begin{bmatrix} 5 & 900 \end{bmatrix}$ cells $\approx \begin{bmatrix} 74 & 13,300 \end{bmatrix}$ km^2
q	90%
A_0	1%
κ	1 mm
f_1, f_2	0.5, 0.3
$S_{\min,1}, S_{\min,2}$	0.3, 0.3
$V_{\max}$	77 km/hr
λ	2.5
R	770 km
ε	± 60 degrees

4.3.3. Storm Tracks

Figure 4-8 shows a sample complete storm track from October 5 at 10:00 p.m. to October 6 at 5:00 a.m. for a storm feature that intersects Brazos County. The 90% quantile level of the storm feature is shown along with the locations and track of the storm feature centroid. As seen in the figure, the storm size decreases over time and storm features clearly overlap from image to image. Although the storm lasts for seven hours, its duration in Brazos County is only one hour in most places. Its average speed over Brazos County is 35 km/hr and a total area of 1,400 km^2 within the county is affected.

We notice also in Figure 4-8 that the overall weighted centroid is used to calculate the storm feature velocity; whereas, each storm feature can be made up of more than one Gaussian component. Each component may move at a different velocity relative to the storm feature centroid. A natural extension of our tracking algorithm could track individual components instead of complete storm features.

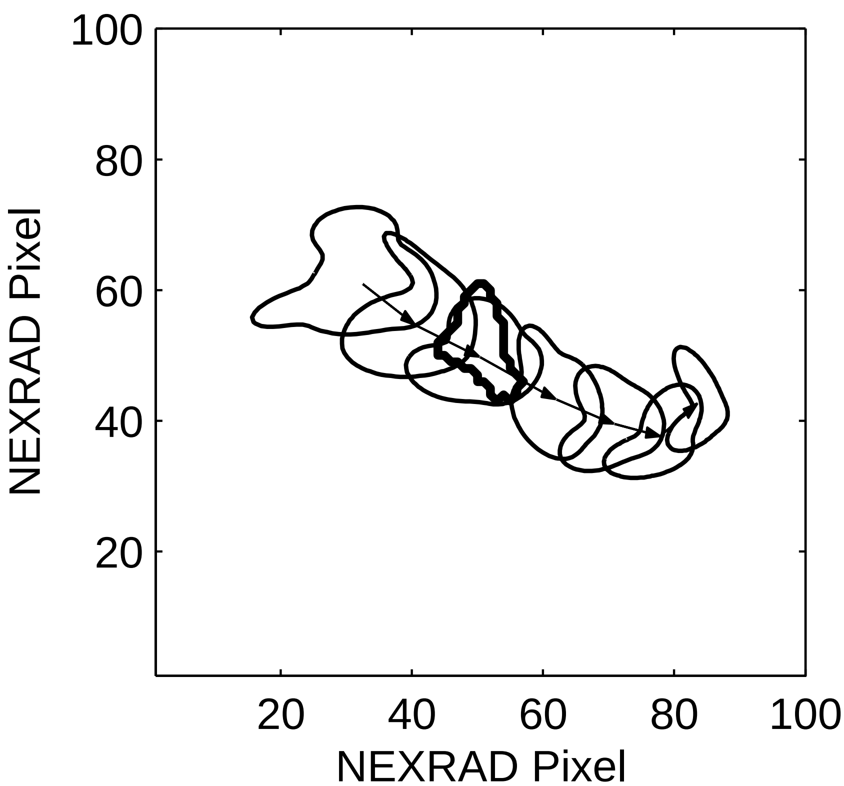


Figure 4-8. Storm feature track over Brazos County in Texas.

This more detailed approach would give additional velocity information within storm features. It would also make the dynamics of storm merging and splitting clearer. However, this higher level of detail is not implemented here because it would add a significant computational cost.

4.3.4. Storm Statistics over Brazos County

A total of 312 storm features intersect Brazos County in 2003. 167 of these storm features last for more than one hour so that they can be tracked and assigned a velocity. The remaining 145 storm features last for less than an hour, are recorded in only one image, and cannot be tracked. As storm features pass over Brazos County, we record their hourly speed, direction, residence time, average size, and average coverage. The residence time is the total time that a storm feature intersects Brazos County based on the number of one-hour NEXRAD images in which the storm feature is detected. This residence time may not coincide with the actual duration at a given point in Brazos County or with the total duration of the storm feature due to its track out-side of Brazos County. Table 4-2 summarizes these data for all tracks of storm features.

Figures 4-9 shows the discrete probability density function of the storm feature speed and Figure 4-10 shows a scatter plot of the average speed, direction of propagation, and size of storm features subdivided into the periods May-August and September-April for the 167 storm feature tracks that intersect Brazos County. The average speed for all storm features was 37 km/hr, with more than 75% of the storm features having velocities between 20 and 60 km/hr. For storms occurring between May through August, there is no apparent preferred direction in the data.

Table 4-2. Summaries of the characteristics of the storm features that interest Brazos County for the application in 2003.

Storm feature characteristics	N	Mean	Median	Std	Min	Max
Speed (km/hr)	167	37.42	38.40	15.08	1.63	76.24
Direction (degree)	167	18.73	28.01	64.44	-148.64	177.54
Residence time (hr)	312	1.33	1	0.61	1	5
Average size (km^2)	312	2,369.30	1,472.80	2,563.50	99.63	17,515
Area coverage of storm features (%)	312	28.56	17.41	27.78	1.77	100

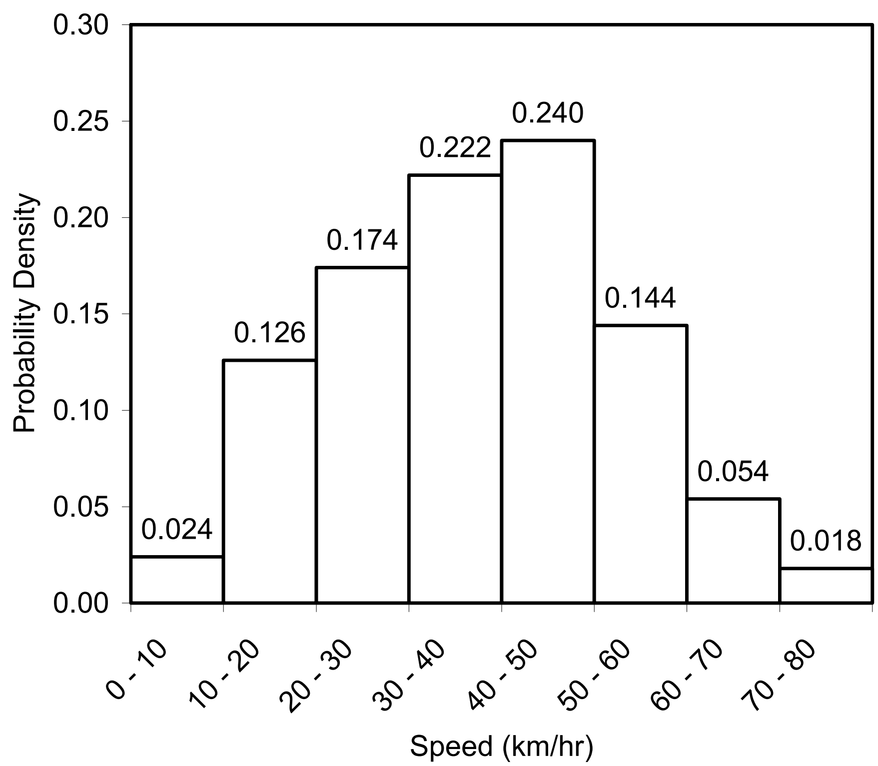


Figure 4-9. Probability density function of storm feature speed for storm features intersecting Brazos County in 2003.

For storms occurring between September through April, there is a clear preference in the data for storm tracks to propagate in the north easterly direction (quadrant I in Figure 4-10). Since the orientation of weather fronts in this area during this period is also in the north-easterly direction, theses results indicate that these storm features propagate along lines of constant pressure (geostrophic flow). Additionally, for the northern hemisphere, the geostrophic flow direction has the low pressure on the left; thus, according to the data set for 2003, most storms occur in the Brazos County when a low pressure front moves toward the southeast.

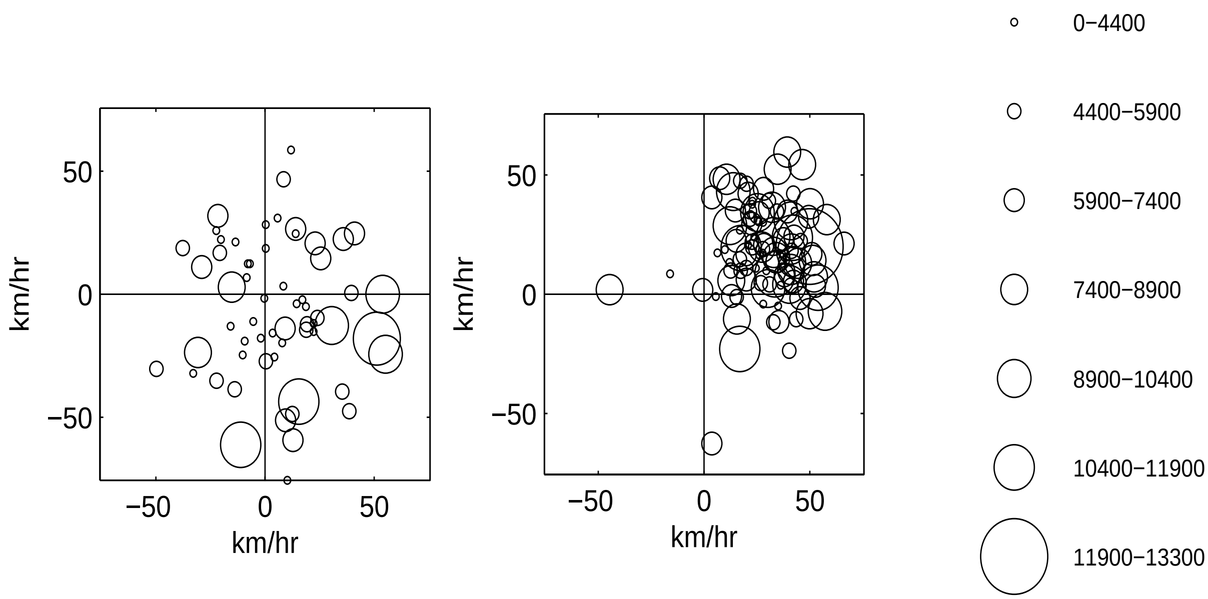


Figure 4-10. Velocity vectors and storm feature sizes (km^2) for storm features intersecting Brazos County in 2003; left: storms between May and August, right: storms between September and April. The circle centroids represents storm feature displacement from (0,0) in one hour; the circle diameters represents storm feature size.

The distribution of storm feature sizes by season is shown in Figures 4-11 (May through August) and 4-12 (September through April). 83% of storm features occurring in the summer (refer to Figure 4-11) are less than 4,000 km^2, and 60% are less than 2,000 km^2, and a median size of 1,600 km^2. This is consistent with the assumption that most summer storms are convective in nature with storm features associated with thunderstorm cells. By contrast, the winter storms (refer to Figure 4-12) have 61% of storm features with average areas less than 4,000 km^2 and 29% less than 2,000 km^2 and a median size of 3,200 km^2. These larger storm feature sizes are consistent with the fact that they are associated with frontal storm systems.

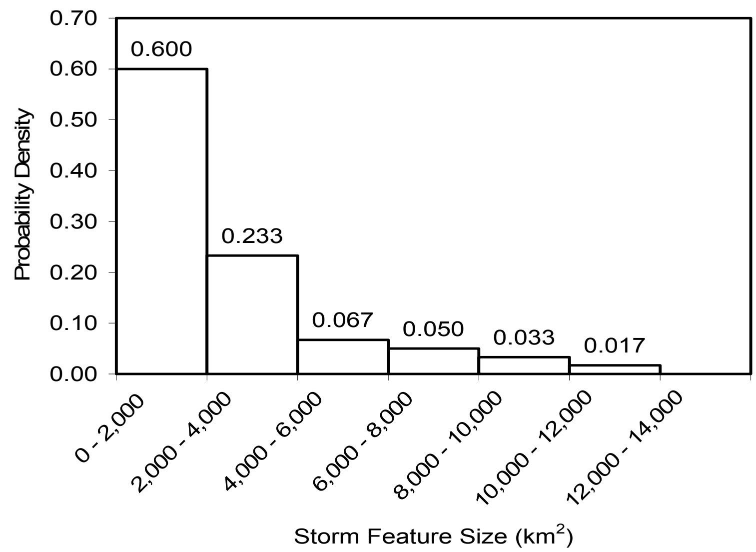


Figure 4-11. Probability density function of storm feature size for May through August for storm features intersecting Brazos County in 2003.

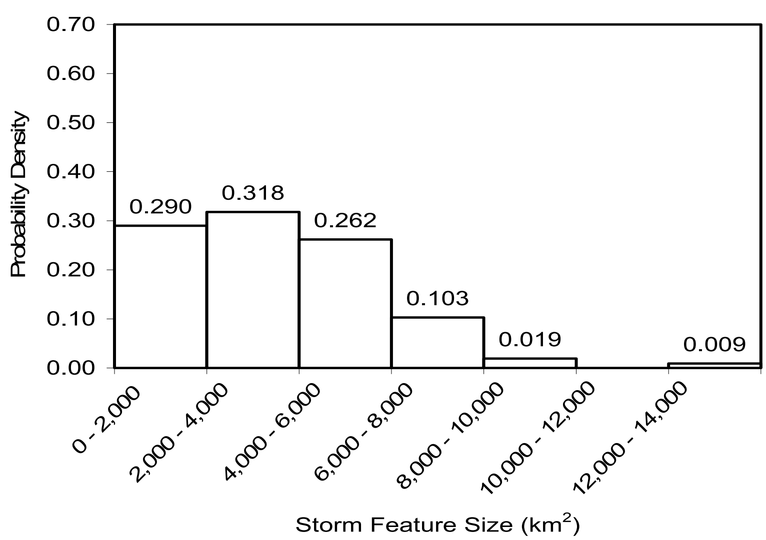


Figure 4-12. Probability density function of storm feature size for September through April for storm features intersecting Brazos County in 2003.

Figure 4-13 shows the discrete probability density function of storm residence time in Brazos County. These values reflect the impact of the average storm speed and propagation direction. Because most storm features move along the short axis of the county (approximately 35 km wide) at a speed of nearly 40 km/hr, 73% of storms have a duration of one hour or less. In addition, since the long axis of the county is approximately 70 km long, 99.4% of the tracked storm features have a duration of 3 hours or less in Brazos County.

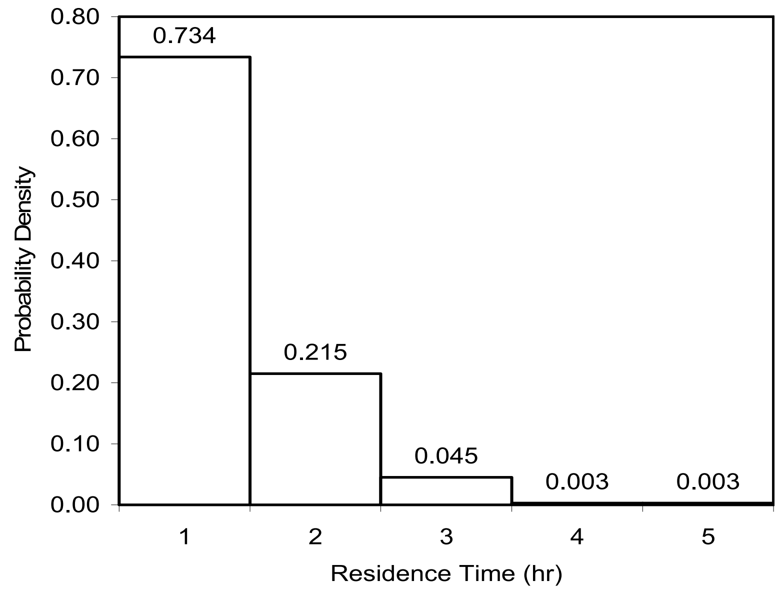


Figure 4-13. Probability density function of storm feature residence time over Brazos County in 2003.

The discrete probability density function of percent areal coverage in Brazos County for tracks of storm features is presented in Figure 4-14. Based on the data in the figure, it is rare for the whole county to receive rainfall from a single storm feature. In fact, only 9.5% of the storm features cover more than 80% of the county. This is despite the fact that most storm features are larger than Brazos County (refer to Figure 4-11 and 4-12). Thus, for Brazos County it can be seen that the assumption that precipitation intensity is uniform over the area or that a storm feature passes through the centroid of the area may be inadequate.

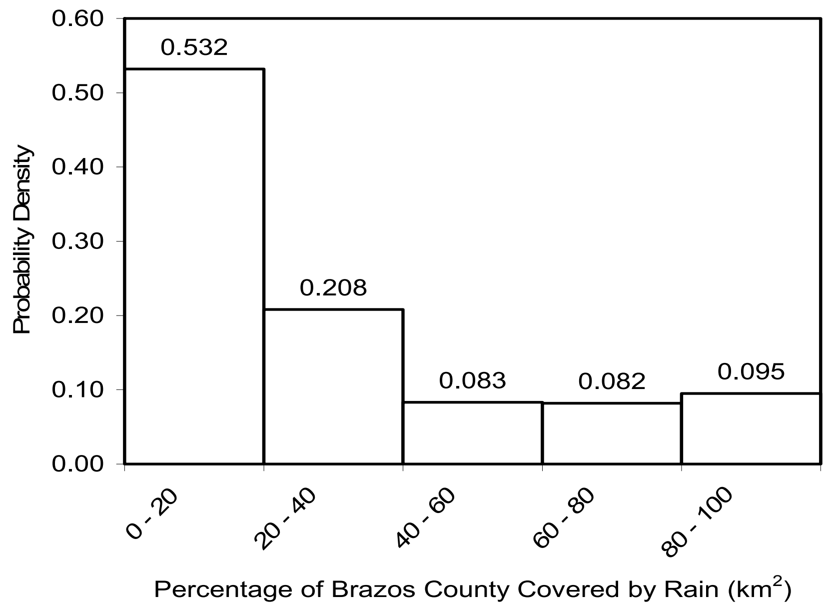


Figure 4-14. Probability density function of the percent area of Brazos County receiving rainfall due to storm features intersecting Brazos County in 2003.

4.4. SUMMARY AND CONCLUSIONS

In this dissertation, we develop a method for identification and tracking of storm features using the hourly-accumulated MPE NEXRAD precipitation data. Application of this method to a specified area allows the determination of the dynamic characteristics of rainfall, such as the storm speed and direction, size, and coverage over a specified area. The method uses image processing and Gaussian mixture models to segment areas of high precipitation and fit their intensity values. Difficulties that stem from aggregation in the one-hour data are overcome by tracking storm features that overlap and by using a new set of inverse cost functions. Given that the NEXRAD data is publicly available, the method can be applied over any area in the United States. The size of the selected area, though, would have a direct effect on the computational cost.

The method is applied to the 1,500 km^2 Brazos County in Texas using the year 2003 NEXRAD data. Based on the dataset used, it was found that as a general trend storm features propagate at around 37 km/hr in the north-east direction. Specifically, more than 75% of the storm features propagate with velocities between 20 km/hr and 60 km/hr, and 61% in directions that range from east to north. Moreover, despite lasting for several hours, storm features pass over at least part of Brazos County and reside in it for less than three hours. In fact, 73% of the storm features reside in the Brazos County for less than one hour. It was also observed that precipitation is very unevenly distributed and that it typically affects only part of the County. In particular, only 9% of the storms features affect more than 80% of its area.

Overall, it was observed that for larger areas, the assumption of uniform rainfall distribution may be inadequate, confirming that hydrologic simulations based on dynamic storms may perform better than those based on uniform precipitation. The method presented here allows the user to better understand the precipitation patterns in any given area of the United States, and yields parameters that describe storm dynamic characteristics. These parameters can then be used in the definition of synthetic dynamic storms for hydrologic modeling.

Further research by the authors will focus on incorporating the intensity of rainfall into the inverse cost functions, and on evaluating the sensitivity of hydrologic simulation models to the use of dynamic as opposed to static storms.

5. SUMMARY

5.1. AREAL REDUCTION FACTORS

It was found that variability of the ARFs is explained in part by seasonality, regionality, and precipitation depth. The seasonality of rainfall, in turn, is related to the governing atmospheric process that produces rainfall of convective or stratiform type. Thus, a sinusoidal regression of the ARFs is used to break the year into two seasons. Different types of storms, however, do occur in the same season and it is not possible to associate one type of storm with one season and vice versa. Therefore, further research is needed to identify storm types from the NEXRAD data and explain the variability of ARF values based on storm type rather than season. The issue of regionality is similar to that of seasonality. The dependence of the ARF values on geographic region is related to the presence of specific types of storms. Therefore, again, ARFs are not really associated with season or region, but with the storm type.

In this study, ARFs were calculated using 18,531 storms of different durations that took place in different seasons and regions of Texas. The calculation of ARFs for a specific duration requires the identification of the storm center, which is the NEXRAD pixel with the maximum accumulated depth within its vicinity. Since the precipitation depth was not considered for identifying the storm, the locations of storms will be distributed across and within the regions of Texas. It is likely, though, that the application of ARFs derived for a region does not account for specific atmospheric processes in the location that make ARF values higher or lower. It is considered that further research is necessary to account for the spatial variability of ARFs at levels finer than the regions used here, which could be identified using clustering analysis. This type of analysis, however, would require many years of NEXRAD data in order to obtain statistically significant results. These data are still not available.

5.2. RAINFALL DISAGGREGATION

The rainfall disaggregation model presented in Section 3 was able to successfully reproduce zero rainfall probabilities, but variance of the hourly precipitation were systematically under-predicted, and lag one-hour autocorrelation coefficients were over-predicted for lower values and under-predicted for higher values. As in any disaggregation method, the storm starting times are unknown and several options can be used to estimate them. In this dissertation, storm starting times are sampled from a storm database, leaving the option of duration overlap. It is believed that this overlap may lead to the biases in the variance and one-hour autocorrelation coefficients mentioned above.

The storm starting time is governed in part by the physical atmospheric processes in the region. For instance, it has been observed that in many parts of the United States, heavy precipitation follows a pronounced diurnal variation (Econopouly et al. 1989). In order to accurately estimate storm starting times and durations, it is necessary to find their patterns in historic precipitation data of the region. In order to reduce the variance of the probability distribution of the starting times and durations, it is necessary to classify storms into groups of similar properties (e.g., based on their types, and/or shape of the hyetograph). The storm starting time and duration would be determined by their probability distribution. However, in this work, a storm was randomly selected from a database of all observed storms, without differentiating by type or hyetograph shape. By classifying storms by type, it is also expected that variance and 1-hr lag autocorrelation will be improved.

5.3. STORM IDENTIFICATION AND TRACKING ALGORITHM

Storm residence time over an area of interest is found to be related to average storm speed and propagation direction. It was found that more than 75% of the storm features have velocities between 20 and 60 km/hr, and 73% of storms have durations of one hour or less 3 hours or less in Brazos County.

Schumacher (2003) studied different types of storms, focusing on mesoscale convective systems (MCSs), which produce extreme rainfall totals in central and eastern United States. He found that most storms producing extreme rainfall events are associated with MCSs during the summer and with synoptically-forced systems in spring and autumn. Over half of the extreme events were found to be associated with MCSs. A MCS forms trailing stratiform and moves parallel to the convective line. A series of storms associated with MCS can pass over a given area consecutively and lead to extreme accumulation of rain depth over an extended period of time. In order to obtain valuable information on the spatial and dynamic patterns of MCSs over an area, it is necessary to identify and track a MCSs objectively, and automated algorithms are necessary. The extended algorithm for storm feature identification and tracking will provide better understand the MCSs patterns in any given area of the United States, and parameters can be obtained from statistical analysis of the extracted information of MCSs. These parameters can then be used in the definition of synthetic dynamic storms for hydrologic modeling.

REFERENCES

Adrian, R. J. (1991). "Particle-imaging techniques for experimental fluid-mechanics." *Annual Review of Fluid Mechanics*, 23, 261-304.

Alley, W. M. (1981). "Estimation of impervious-area washoff parameters." *Water Resources Research*, 17(4), 1161-1166.

Alley, W. M., and Smith, P. E. (1981). "Estimation of accumulation parameters for urban runoff quality modeling." *Water Resources Research*, 17(6), 1657-1664.

Asquith, W. H., and Famiglietti, J. S. (2000), "Precipitation areal-reduction factor estimation using an annual-maxima centered approach." *Journal of Hydrology,* 230 (1-2), 55-69.

Bacchi, B., and Ranzi, R. (1996), "On the derivation of the areal reduction factor of storms" *Atmospheric Research*, 42, 123–135.

Bedient B. P., Holder, A., Benavides, J., and Vieux, B. E. (2003), "Radar-based flood warning system applied to tropical storm Allison." *Journal of Hydrologic Engineering*, 8 (6), 308-318.

Bedient B. P., Hobilt , B. C., Gladwell, D. C., and Vieux, B. E. (2000), "NEXRAD radar for flood prediction in Houston.", *Journal of Hydrologic Engineering*, 5(3), 269-277.

Bellin, A., and Rubin, Y. (1996). "HYDRO_GEN: A spatially distributed random field generator for correlated properties." *Stochastic Hydrology and Hydraulics*, 10(4), 253-278.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics – Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, Inc., New York.

Berenguer, M., Corral, C., Sanchez-Diezma, R., and Sempere-Torres, D. (2005). "Hydrological validation of a radar-based nowcasting technique." *Journal of Hydrometeorology*, 6(4), 532-549.

Blimes, J. A. (1998). A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. TR-97-021, International Computer Science Institute, University of California, Berkeley.

Bo, Z., Islam, S., and Eltahir, E. A. B. (1994). "Aggregation-disaggregation properties of a stochastic rainfall model." *Water Resource Research*, 30(12), 3423-3435.

Bonner, W. D. (1998). *Future of the National Weather Service Cooperative Observer Network.* National Research Council, National Academy Press, Washington, DC.

Bowler, N. E. H., Pierce, C. E., and Seed, A. (2004). "Development of a precipitation nowcasting algorithm based upon optical flow techniques." *Journal of Hydrology,* 288(1-2), 74-91.

Box, G., and Cox, D. (1964). "An analysis of transformations." *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211-252.

Bremaud, P. J., and Pointin, Y. B. (1993). "Forecasting heavy rainfall from rain cell motion using radar data." *Journal of Hydrology*, 142(1-4), 373-389.

Brown, R. A., and Lewis, J. M. (2005). “Path to NEXRAD - Doppler radar development at the National Severe Storms Laboratory.” *Bulletin of the American Meteorological Society*, 86(10), 1459.

Chow, V. T. (1964). *Handbook of Applied Hydrology*, McGraw Hill, New York.

Cowpertwait, P. S. P. (1994). "A generalized point process model for rainfall." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 447(1929), 23-37.

Cowpertwait, P. S. P. (1995). "A generalized spatial-temporal model of rainfall based on a clustered point process." *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*, 450(1938), 163-175.

Cowpertwait, P. S. P. (1998). "A poisson-cluster model of rainfall: high-order moments and extreme values." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 454(1971), 885-898.

Cowpertwait, P. S. P., O’Connell, P. E., Metcalfe, A. V., and Mawdsley, J. A. (1996a). "Stochastic point process modeling of rainfall, II, Regionalization and disaggregation." *Journal of Hydrology*, 175, 47-65.

Cowpertwait, P. S. P., O’Connell, P. E., Metcalfe, A. V., and Mawdsley, J. A. (1996b). "Stochastic point process modelling of rainfall .1. Single-site fitting and validation." *Journal of Hydrology*, 175(1-4), 17-46.

Cox, D. R., and Isham, V. (1988). "A simple spatial-temporal model of rainfall." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 415(1849), 317-328.

Crane, R. K. (1979). "Automatic cell detection and tracking." *IEEE Transactions on Geoscience and Remote Sensing*, 17(4), 250-262.

Crum, T. D., Alberty, R. L., and Burgess, D. W. (1993). “Recording, archiving, and using WSR-88D data.” *Bulletin of the American Meteorological Society*, 74(4), 645-653.

Crum, T. D., and Alberty, R. L. (1993). “The WSR-88D and the WSR-88D operational support facility.” *Bulletin of the American Meteorological Society*, 74(9), 1669-1687.

De Lannoy, G. J. M., Verhoest, N. E. C., and De Troch, F. P. (2005). "Characteristics of rainstorms over a temperate region derived from multiple time series of weather radar images." *Journal of Hydrology*, 307(1-4), 126-144.

De Lima, J. L., and Singh, V. P. (2002). “The influence of the pattern of moving rainstorms on overland flow.” *Advances in Water Resources*, 25(7), 817-828(12)

De Michele C., Kottegoda, N. T., and Rosso, R. (2001). “The derivation of areal reduction factor of storm rainfall from its scaling properties.” *Water Resources Research*, 37(12), 3247-3252.

Dell’Acqua, F., and Gamba, P. (2002). "Rain pattern tracking by means of COTREC and modal matching." *Optical Engineering*, 41(2), 278-286.

Dixon, M., and Wiener, G. (1993). "TITAN - thunderstorm identification, tracking, analysis, and nowcasting - A radar-based methodology." *Journal of Atmospheric and Oceanic Technology*, 10(6), 785-797.

Donigian, A. S., Jr., Imhoff, J. C., Bicknell, B. R., and Kittle, J. L., Jr. (1984). "Application Guide for Hydrological Simulation Program-Fortran (HSPF)." *Environmental Research Laboratory*, Office of Research and Development, Athens, GA.

Durrans S. R, Julian, L. T., and Yekta, M. (2002). “Estimation of depth-area relationships using radar-rainfall data.” *Journal of Hydrologic Engineering,* 7 (5), 356 – 367.

Eagleson, P. S. (1970). *Dynamic Hydrology*, McGraw Hill, New York.

Eagleson, P. S. (1972). “Dynamics of flood frequency.” *Water Resources Research*, 8(4), 878-898.

EarthInfo, I. (2001). "NCDC hourly precipitation." Environmental Databases, EarthInfo, Inc., Boulder, CO.

Econopouly, T. W., Davis, D. R., and Woolhiser, D. A. (1990). "Parameter transferability for a daily rainfall disaggregation model." *Journal of Hydrology*, 118, 209-228.

Einfalt, T., Denoeux, T., and Jacquet, G. (1990). "A radar rainfall forecasting method designed for hydrological purposes." *Journal of Hydrology*, 114(3-4), 229-244.

Fraley, C., and Raftery, A. E. (1998). "How many clusters? Which clustering method? Answers via model-based cluster analysis." *Computer Journal*, 41(8), 578-588.

Frederick, R. H., Myers, V. A., and Auciello, E. P. (1977). “Storm depth-area relations from digitized radar returns.” *Water Resources Research,* 13(3), 675-679.

Fulton, R. A., Breidenbach, J. P., Seo, D. J., Miller, D. A., and O'Bannon, T. (1998). "The WSR-88D rainfall algorithm." *Weather and Forecasting*, 13(2), 377-395.

Gebremichael, M., and Krajewski W. F. (2004). “Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets.” *Journal of Applied Meteorology* 43, 1180-1199.

Giannoni F., Smith, J. A., Zhang, Y., and Roth, G. (2003). “Hydrologic modeling of extreme floods using radar rainfall estimates.” *Advances in Water Resources,* 26, 195-203.

Glasbey, C. A., Cooper, G., and McGechan, M. B. (1995). "Disaggregation of daily rainfall by conditional simulation from a point-process model." *Journal of Hydrology*, 165(1-4), 1-9.

Goodrich D. C., Faur`es J-M, Woolhiser D. A., Lane L. J., and Sorooshian, S. (1995). "Measurement and analysis of small-scale convective storm rainfall variability." *Journal of Hydrology*, 173, 283–308.

Grecu, M., and Krajewski, W. F. (2000). "A large-sample investigation of statistical procedures for radar-based short-term quantitative precipitation forecasting." *Journal of Hydrology*, 239(1-4), 69-84.

Güntner, A., Olsson, J., Calver, A., and Gannon, B. (2001). "Cascade-based disaggregation of continuous rainfall time-series: the influence of climate." *Hydrology and Earth System Sciences*, 5(2), 145-164

Gupta, V. K., and Waymire, E. C. (1993). "A statistical-analysis of mesoscale rainfall as a random cascade." *Journal of Applied Meteorology*, 32(2), 251-267.

Gutierrez-Magness, A. L., and McCuen, R. H. (2004). "Accuracy evaluation of rainfall disaggregation methods." *Journal of Hydrologic Engineering*, 9(2), 71-78.

Gyasi-Agyei, Y. (2005). "Stochastic disaggregation of daily rainfall into one-hour time scale." *Journal of Hydrology*, 309, 178-190.

Wheater, H. S., Isham, V. S., Cox, D. R., Chandler, R. E., Kakou, A., Northrop, P. J., Oh, L., Onaf, C., and Rodriguez-Iturbe, I. (2000). "Spatial-temporal rainfall fields: modelling and statistical aspects." *Hydrology and Earth System Sciences,* 4, 581-601

Handwerker, J. (2002). "Cell tracking with TRACE3D - a new algorithm." *Atmospheric Research*, 61(1), 15-34.

Hershenhorn, J., and Woolhiser, D. A. (1987). "Disaggregation of daily rainfall." *Journal of Hydrology*, 95(3-4), 299-322.

Hingray, B., and Haha, M. B. (2005). "Statistical performances of various deterministic and stochastic models for rainfall series disaggregation." *Atmospheric. Research*, 77, 152-175.

Houze, R. A., Brodzik, S., Schumacher, C., Yuter, S. E., and Williams, C. R. (2004). "Uncertainties in oceanic radar rain maps at Kwajalein and implications for satellite validation." *Journal of Applied Meteorology*, 43, 1114-1132.

Islam, S., Entekhabi, D., and Bras, R. L. (1990). "Parameter estimation and sensitivity analysis for the Modified Bartlett-Lewis Rectangular Pulses Model of rainfall." *Journal of Geophyical Research.*, 95(D3), 2093-2100.

Jain, A. K., Duin, R. P. W., and Mao, J. (2000). "Statistical pattern recognition: A review." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(1), 4-37.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, E. D., Stumpf, G. J., Eilts, M. D., and Thomas, K. W. (1998). "The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm." *Weather and Forecasting*, 13(2), 263-276.

Johnson, K., Hoef, J. M. V., Krivoruchko, K., and Lucas, N. (2001). *Using ArcGIS Geostatistical Analyst.* ESRI Press, Redlands, CA.

Klazura, G. E., and Imy, D. A. (1993). "A description of the initial set of analysis products available from the NEXRAD WSR-88D system." *Bulletin of the American Meteorological Society*, 74(7), 1293-1311.

Kondragunta, C., Kitzmiller, D., Seo, D. J., and Shrestha, K. (2005). "Objective integration of satellite, rain gauge, and radar precipitation estimates in the multisensor precipitation estimator algorithm." 85th American Meteorological Society Annual Meeting, 19th Conf. Hydrology, San Diego, CA.

Koutsoyiannis, D., and Onof, C. (2001). "Rainfall disaggregation using adjusting procedures on a Poisson cluster model." *Journal of Hydrology*, 246, 109-122.

Koutsoyiannis, D., Onof, C., and Wheater, H. S. (2003). "Multivariate rainfall disaggregation at a fine timescale." *Water Resources Research*, 39(7), 1173.

Krajewski, W. F., Raghavan, R., and Chandrasekar, V. (1993). "Physically based simulation of radar rainfall data using a space-time rainfall model." *Journal of Applied Meteorology*, 32(2), 268-283.

Krajewski, W. F., and Smith, J. A. (2002), "Radar hydrology: rainfall estimation." *Advances in Water Resources,* 25(2002), 1387–1394.

Lakshmanan, V., Rabin, R., and DeBrunner, V. (2003). "Multiscale storm identification and forecast." *Atmospheric Research*, 67-8, 367-380.

Le Cam, L. (1961). "A stochastic description of precipitation." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, 165-186.

Legates, D. R., and McCabe, G. J. (1999). "Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation." *Water Resources Research*, 35(1), 233-241.

Li, L., Schmid, W., and Joss, J. (1995). "Nowcasting of motion and growth of precipitation with radar over a complex orography." *Journal of Applied Meteorology*, 34(6), 1286-1300.

Likas, A., Vlassis, N., and Verbeek, J. (2003). "The global k-means clustering algorithm." *Pattern Recognition*, 36, 451-461.

Lovejoy, S. (1982). "Area-perimeter relation for rain and cloud areas." *Science*, 216, 185-187.

Lovejoy, S., and Schertzer, D. (1986). "Scale-invariance, symmetries, fractals, and stochastic simulations of atmospheric phenomena." *Bulletin of the American Meteorological Society*, 67(1), 21-32.

Lovejoy, S., and Schertzer, D. (2006). "Multifractals, cloud radiances and rain." *Journal of Hydrology*, 322(1-4), 59-88.

Lovejoy, S., Schertzer, D., and Tsonis, A. A. (1987). “Functional box-counting and multiple dimensions in rain.” *Science*, 235, 1036-1038.

Marani, M., Grossi, G., Napolitano, F., and Wallace, M. (1997). "Forcing, intermittency, and land surface hydrologic partitioning." *Water Resources Research*, 33(1), 167-175.

Margulis, S. A., and Entekhabi, D. (2001). "Temporal disaggregation of satellite-derived monthly precipitation estimates and the resulting propagation of error in partitioning of water at the land surface." *Hydrology Earth System Sciences*, 5(1), 27-38.

McCollum, J. R., Krajewski, W. F., Ferraro, R. R., and Ba, M. B. (2002). “Evaluation of biases of satellite rainfall estimation algorithms over the continental United States.” *Journal of Applied Meteorology*, 41, 1065-1080.

Mecklenburg, S., Bell, V. A., Moore, R. J., and Joss, J. (2000a). "Interfacing an enhanced radar echo tracking algorithm with a rainfall-runoff model for real-time flood forecasting." *Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere*, 25(10-12), 1329-1333.

Mecklenburg, S., Joss, J., and Schmid, W. (2000b). "Improving the nowcasting of precipitation in an Alpine region with an enhanced radar echo tracking algorithm." *Journal of Hydrology*, 239(1-4), 46-68.

Molnar, P., and Burlando, P. (2005). "Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model." *Atmospheric Research*, 77, 137-151.

Moon, J., Srinivasan, R., and Jacobs, J. H. (2004). “Stream flow estimation using spatially distributed rainfall in the Trinity River basin.” *Transactions of the American Society of Agricultural and Biological Engineers,* 47(5), 1445-1451.

Morin, E., Goodrich, D. C., Maddox, R. A., Gao, X. G., Gupta, H. V., and Sorooshian, S. (2006). "Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response." *Advances in Water Resources*, 29(6), 843-860.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models, Part I-A discussion of principles." *Journal of Hydrology*, 10, 282-290.

National Weather Service. (2006). “NEXRAD stage III (MPE) distributed precipitation data.” Radar Operations Center, National Weather Service, <http://www.roc.noaa.gov/> (Aug. 10, 2006).

National Climate Data Center. (2003). "Data documentation for data set 3240 (DSI-3240) hourly precipitation data." *National Climate Data Center (NCDC)*, Asheville, NC.

National Weather Service. (1980). “A methodology for point-to-area rainfall frequency ratios.” *Technical Report 24*, Office of Hydrology, National Weather Service, Silver Spring, MD.

National Weather Service. (2002). “About stage III data.” National Weather Service, <http://www.nws.noaa.gov/oh/hrl/dmip/stageiii_info.htm> (June 1, 2005).

National Weather Service. (2005a). “Radar operations center.”, National Weather Service < http://www.roc.noaa.gov/ > (Oct. 25, 2005).

National Weather Service. (2005b). “Archive of river forecast center operational NEXRAD data, National Weather Service.” <http://dipper.nws.noaa.gov/hdsb/data/nexrad/nexrad.html> (Oct. 26, 2005).

Nelson, B. R., Kim, D., Seo, D.-J., and Bates, J. (2006). "Multi-sensor precipitation reanalysis." 86th American Meteorological Society Annual Meeting, 20th Conf. Hydrology, Atlanta, GA.

Northrop, P. (1998). "A clustered spatial-temporal model of rainfall." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 454(1975), 1875-1888.

Ogden F. L., Richardson J. R., and Julien P. Y. (1995). “Similarity in catchment response. 2. Moving storms.” *Water Resource Research,* 31(6), 1543–7.

Ogden, F. L., and Julien, P. Y. (1993). “Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales.” *Water Resources Research*, 29 (8), 2589-2597.

Ogden, F. L., and Julien, P.Y. (1994). “Runoff model sensitivity to radar rainfall resolution.” *Journal of Hydrology,* 158 (1-2), 1-18.

Ogden, F. L., Sharif, H. O., Senarath, S. U. S., Smith, J. A., Baeck M. L., and Richardson, J. R. (2000). “Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997.” *Journal of Hydrology*, 228(1-2), 82-100.

Olsson, J. (1998). "Evaluation of a scaling cascade model for temporal rainfall disaggregation." *Hydrology and Earth System Sciences*, 2(1), 19-30.

Olsson, J., and Berndtsson, R. (1998). "Temporal rainfall disaggregation based on scaling properties." *Water Science and Technology*, 37(11), 73-79.

Omolayo, A.S. (1993), “On the transposition of areal reduction factors for rainfall frequency estimation.” *Journal of Hydrology*, 145(1993), 191-205.

Onof, C., and Wheater, H. S. (1993). "Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulses Model." *Journal of Hydrology*, 149(1-4), 67-95.

Onof, C., and Wheater, H. S. (1994). "Improvements to the modeling of British rainfall using a modified random parameter Bartlett-Lewis Rectangular Pulse Model." *Journal of Hydrology*, 157(1-4), 177-195.

Ormsbee, L. E. (1989). "Rainfall disaggregation model for continuous hydrologic modeling." *Journal of Hydraulic Engineering*, 115(4), 507-525.

Osborn, H. B., and Lane, L. J. (1981). “Point-area-frequency conversion for summer rainfall in Southeastern Arizona.” *Hydrology and Water Resources of Arizona and the Southwest*, 11, 39-42.

Phillip, B. B., and Wayne, C. H. (1992). *Hydrology and Floodplain Analysis*, Addison-Wesley Publishing Company, Reading, MA.

Raffel, M., Willert, C., and Kompenhans, J. (1998). *Particle Image Velocimetry - A Practical Guide*, Springer Verlag, Berlin.

Redner, R. A., and Walker, H. F. (1984). "Mixture densities, maximum-likelihood and the EM algorithm." *SIAM Review*, 26(2), 195-237.

Reed, S. M., and Maidment, D. R. (1999). "Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling." *Journal of Hydrologic Engineering*, 4(2), 174-182.

Rinehart, R. E., and Garvey, E. T. (1978). "3-dimensional storm motion detection by conventional weather radar." *Nature*, 273(5660), 287-289.

Rodriguez-Iturbe, I., and Mejia, J. M. (1974). “On the transformation of point rainfall to areal rainfall.” *Water Resources Research*, 10(4), 729-735.

Rodriguez-Iturbe, I., Cox, D. R., and Eagleson, P. S. (1986). "Spatial modeling of total storm rainfall." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 403(1824), 27-50.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). "Some models for rainfall based on stochastic point processes." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 410(1839), 269-288.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1988). "A point process model for rainfall - Further developments." *Proceedings of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 417(1853), 283-298.

Rodriguez-Iturbe, I., Marani, M., D'Odorico, P., and Rinaldo, A. (1998). "On space-time scaling of cumulated rainfall fields." *Water Resources Research*, 34(12), 3461-3469.

Schertzer, D., and Lovejoy, S. (1987). "Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes." *Journal of Geophysical Research-Atmospheres*, 92(D8), 9693-9714.

Schumacher, R. S., and Johnson, R. H. (2005). "Organization and environmental properties of extreme-rain-producing mesoscale convective systems." *Monthly Weather Review*, 133(4), 961-976.

Schumacher, R. S., and Johnson, R. H. (2006). "Characteristics of US extreme rain events during 1999-2003." *Weather and Forecasting*, 21(1), 69-85.

Schweppe, F. C. (1973). *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Seed, A.W., Srikanthan, R., and Menabde, M. (1999). “A space and time model for design storm rainfall.” *Journal of Geophysical Research-Atmospheres*, 104(D24), 31623-31630

Seo, D. J., Breidenbach, J. P., and Johnson, E. R. (1999). "Real-time estimation of mean field bias in radar rainfall data." *Journal of Hydrology*, 223(3-4), 131-147.

Siriwardena, L., and Weinmann, P. E. (1996). “Derivation of areal reduction factors for design rainfalls in Victoria.” *Report 96/4*, Cooperative Research Center for Catchment Hydrology < http://www.catchment.crc.org.au/pdfs/technical199604.pdf> (June 2, 2005).

Sivapalan M., and Blöschl, G. (1998). “Transformation of point rainfall to areal rainfall: intensity-duration-frequency curves.” *Journal of Hydrology*, 204, 150–167.

Smith J. A., Baeck, M. L., Zhang, Y., and Doswell, C. (2001). “Extreme rainfall and flooding from supercell thunderstorms.” *Journal of Hydrometeorology*, 2(5), 469–489.

Socolofsky, S. A., Adams, E. E., and Entekhabi, D. (2001). "Disaggregation of daily rainfall for continuous watershed modeling." *Journal Hydrologic Engineering*, 6(4), 300-309.

Socolofsky, S. A., and Adams, E. E. (2003). "Liquid volume fluxes in stratified multiphase plumes." *Journal Hydrologic Engineering*, 129(11), 905-914.

Srikathan, R. (1995). “A review of the methods for estimating areal reduction factors for design rainfalls.” *Report 95/3*, Cooperative Research Center for Catchment Hydrology, <http://www.catchment.crc.org.au/pdfs/technical199503.pdf> (Aug. 2, 2005).

Steinacker, R., Dorninger, M., Wolfelmaier, E., and Krennert, T. (2000). "Automatic tracking of convective cells and cell complexes from lightning and radar data." *Meteorology and Atmospheric Physics*, 72(2-4), 101-110.

Tessier, Y., Lovejoy, S., and Schertzer, D. (1993). "Universal multifractals - theory and observations for rain and clouds." *Journal of Applied Meteorology*, 32(2), 223-250.

Texas Water Development Board. (1967). “The climate and physiography of Texas, Texas Water Development Board.” *Report 53*, 27, <http://rio.twdb.state.tx.us/publications /reports/GroundWaterReports/GWReports/R53/R53.pdf> (Oct. 25, 2005).

Texas Water Development Board. (2006). “TWDB Mapping Website.” *Texas Water Development Board,* < http://www.twdb.state.tx.us/mapping/gisdata.asp. > (Oct. 25, 2005).

Upton, G. J. G. (2000). "Using volumetric radar data to track horizontal and vertical movements of storms." *Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere*, 25(10-12), 1117-1121.

U.S. Geological Survey. (1998). *Extreme precipitation depth for Texas, excluding the Trans-Pecos region*, Water Resources Investigations Report 98-4099, Austin, TX.

Veneziano, D., and Villani, P. (1996). "Identification of rain cells from radar and stochastic modelling of space-time rainfall." *Meccanica*, 31(1), 27-42.

Veneziano, D., Bras, R. L., and Niemann, J. D. (1996). "Nonlinearity and self-similarity of rainfall in time and a stochastic model." *Journal of Geophysical Research-Atmospheres*, 101(D21), 26371-26392.

Vieux B. E., and Vieux, J. E. (2004). "Rainfall accuracy considerations using radar and rain gauge networks for rainfall-runoff monitoring." *Effective Modeling of Urban Water Systems*, Monograph 13.

Vieux B. E., and Bedient, P.B. (2004). "Assessing urban hydrologic prediction accuracy through event reconstruction." *Journal of Hydrology*, 299(3-4), 217-236.

Weather Bureau (1957). "Rainfall intensity-frequency regime – Part 1: The Ohio Valley." *Technical paper No. 29 (TP-29)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1958a). "Rainfall intensity-frequency regime – Part 2: The Southeastern United States." *Technical paper No. 29 (TP-29)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1958b). "Rainfall intensity-frequency regime – Part 3: The Middle Atlantic Region." *Technical paper No. 29 (TP-29)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1959). "Rainfall intensity-frequency regime – Part 4: The Northeastern United States." *Technical paper No. 29 (TP-29)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1960). "Rainfall intensity-frequency regime – Part 5: The Great Lakes Region." *Technical paper No. 29 (TP-29)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1961). "Rainfall frequency atlas of the United States." *Technical paper No. 40 (TP-40)*, U.S. Dept. Of Commerce, Washington, DC.

Weather Bureau (1964). "Two- to ten-day precipitation for return periods of 2 to 100 years in the contiguous United States." *Technical Paper No 49 (TP-49)*, U.S. Dept. Of Commerce,Washington, DC.

Wheater, H. S., Isham, V. S., Cox, D. R., Chandler, R. E., Kakou, A., Northrop, P. J., Oh, L., Onof, C., and Rodriguez-Iturbe, I. (2000). "Spatial-temporal rainfall fields: modelling and statistical aspects." *Hydrology and Earth System Sciences*, 4(4), 581-601.

Willmott, C. J. (1982). "Some comments on the evaluation of model performance." *Bulletin of the American Meteorological Society*, 63(11), 1309-1313.

Wilson, C. B., Valdes J. B., and Rodriguez-Iturbe, I. (1979). "On the influence of the spatial distribution of rainfall on storm runoff." *Water Resources Research,* 15(22)

Wilson, J. W., and Brandes, E. A. (1979). "Radar measurement of rainfall - A summary." *Bulletin of the American Meteorological Society*, 60(9), 1048-1058.

Wilson, J. W., Crook, N. A., Mueller, C. K., Sun, J. Z., and Dixon, M. (1998). "Nowcasting thunderstorms: A Status Report." *Bulletin of the American Meteorological Society*, 79(10), 2079-2099.

Woolhiser, D. A., and Osborn, H. B. (1985). "A stochastic model of thunderstorm rainfall intensities." *Water Resources Research*, 21(4), 511-522.

Young, C. B., Bradley, A. A., Krajewski, W. F., Kruger, A., and Morrisey, M. L. (2000). "Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic Forecasting." *Journal of Hydrometeorology*, 1(3), 241–254.

Zawadzki, I. I. (1973). "Statistical properties of precipitation patterns." *Journal of Applied Meteorology*, 12(3), 459-472.

Zeng , N., Shuttleworth, J. W., and Gash, J. H. C. (2000). "Influence of temporal variability of rainfall on interception loss. Part I. Point analysis." *Journal of Hydrology*, 228, 228-241.

APPENDIX A

This appendix presents the methods used for the derivation of ARFs, the disaggregation model, and storm feature identification and tracking algorithm. This appendix consists of three sections in which each method is described.

The programs used for the methods were SPSS 12.0.1, ArcGIS 9.0, DOS, Cygwin, FORTRAN, C and C++, and MATLAB 7.0. There were different types of functions used in this dissertation. The functions were created in different programming languages: C and C++, FORTRAN, Visual Basic for Application (VBA), MATLAB m-file, and SPSS script language. The use of a function name alone may be confusing to identify which program language is used to create the function. In order to easily identify the program that uses a function, naming conventions were established by putting prefix in the function name. Special characters ($, &, @, #, and %) were used for the prefixes. The different types of functions with a prefix were listed in number as follows. For example, the MATLAB function named "find" is written as &find or find.m in this appendix.

1. $function_name – SPSS script function in a SPSS script file (.SBS).
2. &function_name or function_name.m – MATLAB m-file (.m)
3. @function_name –Visual Basic function in a Visual Basic file (.bas)
4. #function_name – FORTRAN
5. %function_name or function_name.c - C or C++

For other conventions, when a function named "A" calls another function named "B" inside the function "A", the special character "::" is used to express such a relation in this appendix. For example, the expression of A::B indicates that the function B is called by the function A. When a Visual Basic function is created in a module, the calling of the function can be expressed as module_name.function_name using the connector “.”. For example, the expression of A.B indicates that the function named “B” is inside the module named “A”.

A.1. DERIVATION OF ARFS

The section A.1 describes the directory structure storing all the associated files for the statistical analysis, utilities, and the methods applied to derive ARFs.

A.1.1. Directory Structure

All the input, output and script files for the analysis were stored in the root directory c:/research/TXDOT/statistical analysis/ and its subdirectories. Top-level directories were located under the root directory. The top-level directories are considered as a program package that contains all necessary files in the directory structure. Thus, the top-level directories can be moved to any root directory. The directories are listed below.

The top-level directories needed for the statistical analysis

/ancova	Directory associated with ANCOVA test
/scripts	Directory that has SPSS script files (.SBS)
/dataDocs	SPSS data directory that contains SPSS data files (.sav)
/text	Text files (.txt and .dat)
/matlab	Matlab source codes
/2003data	NEXRAD MPE archived data of year 2003 downloaded from website
/2004data	NEXRAD MPE archived data of year 2004 downloaded form website

Subdirectories

The section provides information about some of the subdirectories and their contents.

/dataDocs

*.sav	SPSS data files for statistical analysis
/source	original SPSS data files

/ancova

/dataDocs	SPSS data files for ANCOVA test
/outputDocs	SPSS output files for ANCOVA test

/ancova/cat_rgn

/outputDocs	SPSS output files from ANCOVA test for region
/dataDocs	SPSS data files for ANCOVA test for region
/reports	documents for report

/ancova/cat_ss

/outputDocs	SPSS output files from ANCOVA test for season
/dataDocs	SPSS data files for ANCOVA test for season
/reports	documents for report

/2003/mpe/

.	Archived NEXRAD MPE files (.rar)
/commonfiles	NEXRAD MPE binary files of year 2003

/2004/mpe/

.	Archived NEXRAD MPE files (.rar)
/commonfiles	NEXRAD MPE binary files of year 2004

A.1.2. Genetic Algorithm to Detect the Shape and the Direction of the Storm

Firstly, brief description regarding the radar precipitation data and the way to obtain it through internet resources is provided. Then, the genetic algorithm that detects the shape and the direction of storm is explained. Then, the procedure to run the MATLAB codes that were developed to implement the genetic algorithm will be described.

1. Precipitation Data Acquisition

NEXRAD Precipitation data is provided from Western Gulf River Forecasting Center (WGRFC), which covers part of Texas, Colorado and New Mexico. The approximate area of coverage is shown in Figure A-1.

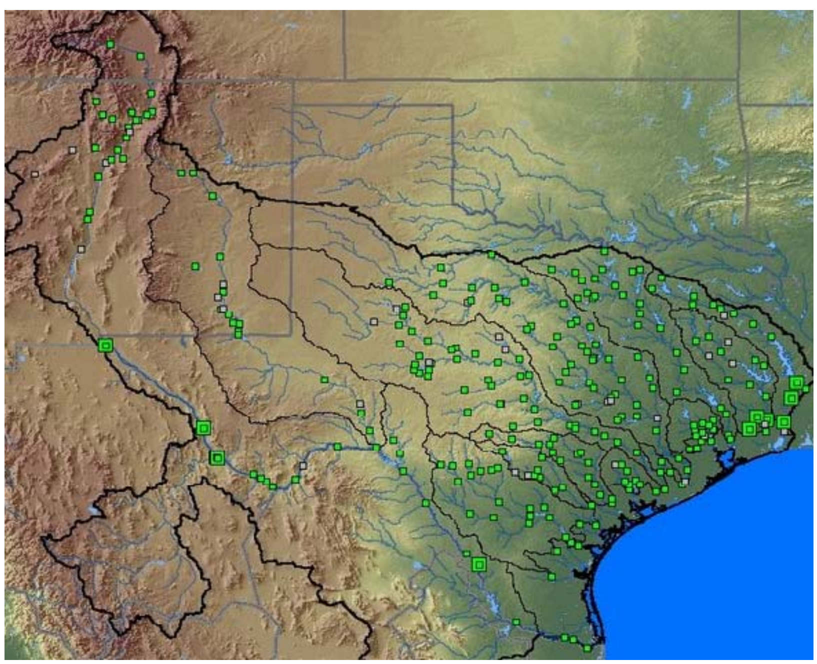


Figure A-1. Coverage of WGRFC.

More specific coverage of the radar precipitation data is shown as a distorted rectangle in Figure 2-1. The rectangle is subdivided into a grid system called "HRAP" with 390 rows and 425 columns. Each cell has precipitation data in the unit of mm and has an area of ~16km^2 depending on the location of the cell in the grid system.

The way the precipitation is recorded using radar system is well described in the publications that can be found in the following internet website, shown in Figure A-2.

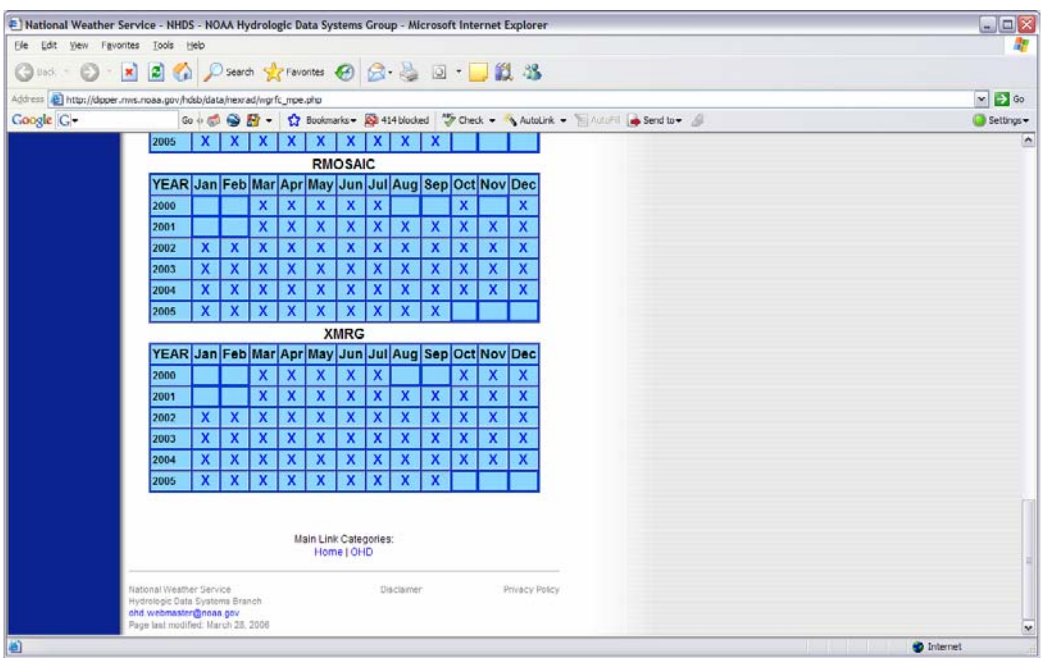
http://www.roc.noaa.gov/app/publications.asp

http://www.ncdc.noaa.gov/oa/radar/radarresources.html

The data can be downloaded from the following website.

http://dipper.nws.noaa.gov/hdsb/data/nexrad/nexrad.html → for entire States

http://dipper.nws.noaa.gov/hdsb/data/nexrad/wgrfc_mpe.php → for WGRFC



2005	X	X	X	X	X	X	X	X	X			

RMOSAIC

YEAR	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000			X	X	X	X	X			X		X
2001			X	X	X	X	X	X	X	X	X	X
2002	X	X	X	X	X	X	X	X	X	X	X	X
2003	X	X	X	X	X	X	X	X	X	X	X	X
2004	X	X	X	X	X	X	X	X	X	X	X	X
2005	X	X	X	X	X	X	X	X	X			

XMRG

YEAR	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000			X	X	X	X	X			X	X	X
2001			X	X	X	X	X	X	X	X	X	X
2002	X	X	X	X	X	X	X	X	X	X	X	X
2003	X	X	X	X	X	X	X	X	X	X	X	X
2004	X	X	X	X	X	X	X	X	X	X	X	X
2005	X	X	X	X	X	X	X	X	X			

Figure A-2. XMRG data.

The downloaded data is a zipped archive that contains the daily precipitation data of the region. You will see a window that similar to Figure A-3. It is recommended to use WinRAR software for zipping and unzipping since it is the most time-efficient software. The software can be downloaded from www.rarlab.com. It is free of license.

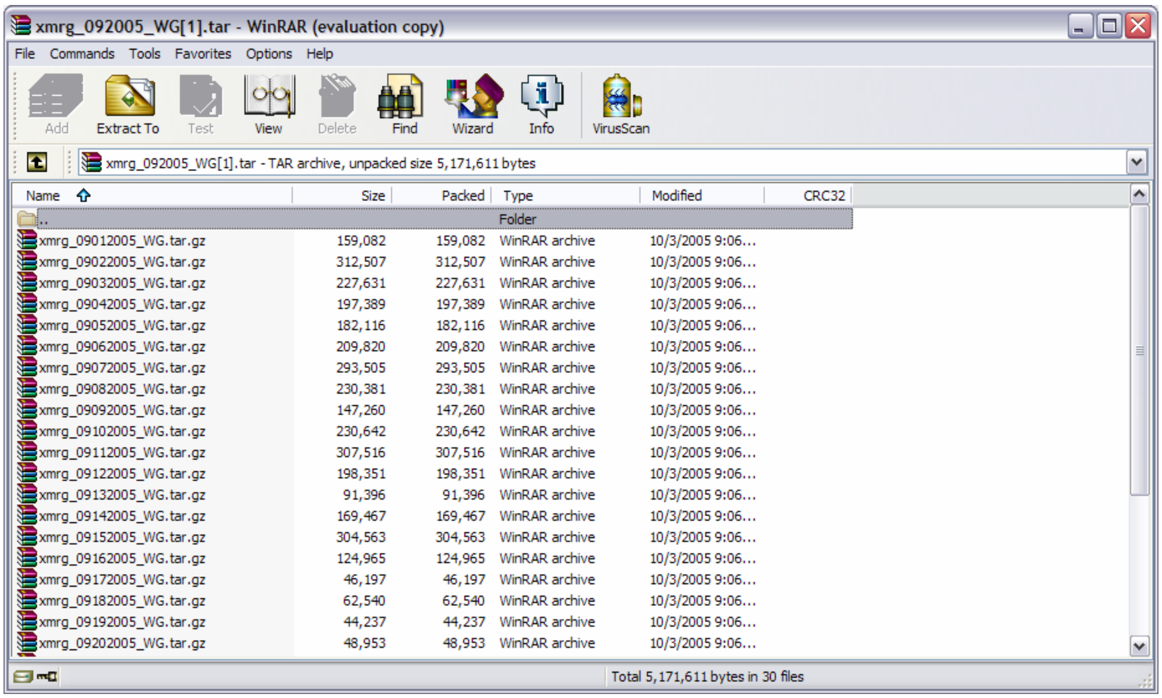


Figure A-3. The *.tar.gz file contains 24 hourly precipitation data.

Once the archived files were unzipped, the unzipped files have the file extension ".tar.gz" which is also an archived file. The files were again unzipped to create the binary files with no file extension, as shown in Figure A-4.

The xmrg file name without any extension is typical example of name of this file. This file is not in ASCII format, so you can't read or process it even though you double click on and open it using applications like notepad or excel. To convert the file into readable format, you have to download a set of c code and compile and run it on the precipitation data you downloaded. This set of c code can be found on the following website. Methodology of decoding this file is also provided in various documents in this website “http://www.nws.noaa.gov /oh/hrl/pps/pps.htm”.

The format of converted precipitation data is as follows;

1 1 0

1 2 0.3

1 3 0.5

…

390 425 0

(Format 1)

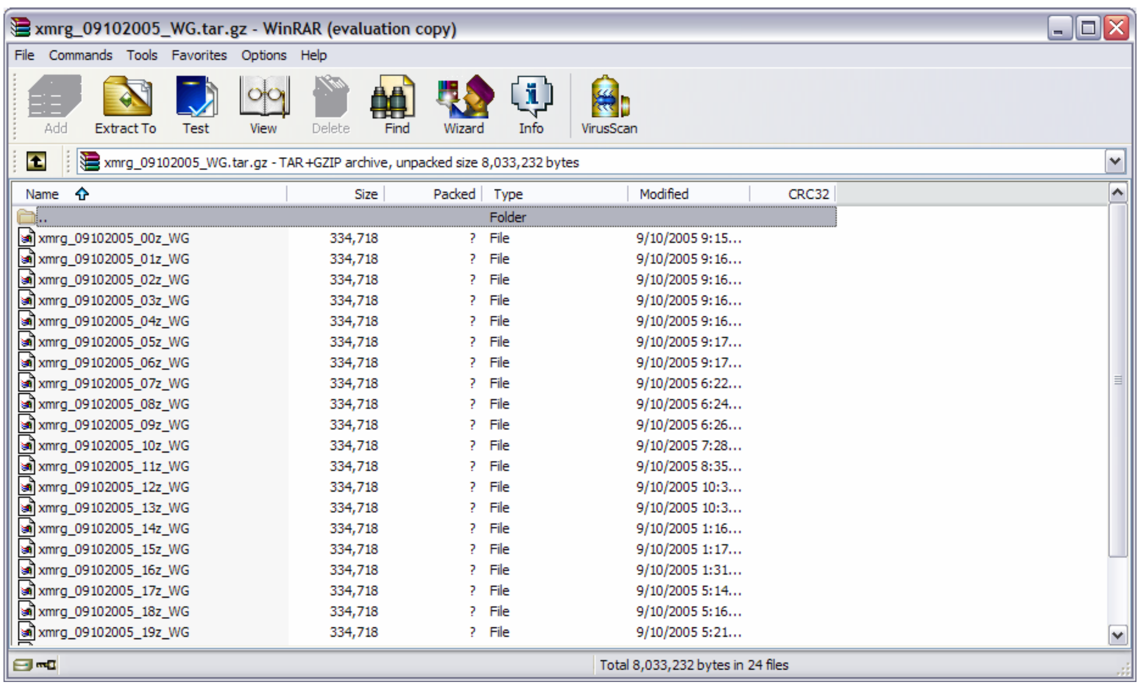


Figure A-4. xmrg_09102005_00z_WG.

The first column indicates the row number, and the second column indicates the column number in the HRAP grid that was mentioned earlier. The third column indicates the depth of the precipitation that corresponds to that row and column. Row and column starts at the upper left-most corner as follows;

(1,1) (1,2) (1,3) (1,4) (1,5) …… (1,425)

(2,1) (2,2) (2,3) (2,4) (2,5) …… (2,425)

……

(390, 1) (390, 2) ………   (390, 425)

(Format 2)

Format I of the precipitation data is extremely inefficient in terms of data processing since we have to read the data all the way from the first row to retrieve the data we want to read. For example, to access the precipitation depth in the row 200 and column 200 of a certain hour of a given day, the program should skip 199×425 + 199 lines of the ASCII file. A MATLAB code was developed to convert the dataset with (Format 1) into (Format 2). A function &mtr_converter does this job. Once the files with format 1 were converted in to format2, the program can access the same data in the aforementioned example with skipping only 200 lines. This code requires as input of a file with the following format:

xmrg_01012004_00z_WG1

xmrg_01012004_01z_WG2

….

xmrg_12312004_23z_WG3

The first column of the file is the name of the ASCII file in "format1". This data conversion process was performed to 2003 and 2004 MPE data and those data were zipped in the file named "2003_mpe_mtr.rar" and "2004_mpe_mtr.rar". No further data processing is necessary if one is only to use 2003 and 2004 MPE data. However, if 2003 and 2004 MPE data were to be used in research, the steps described in this section should be performed to acquire the data. You should get the data that looks as follows:

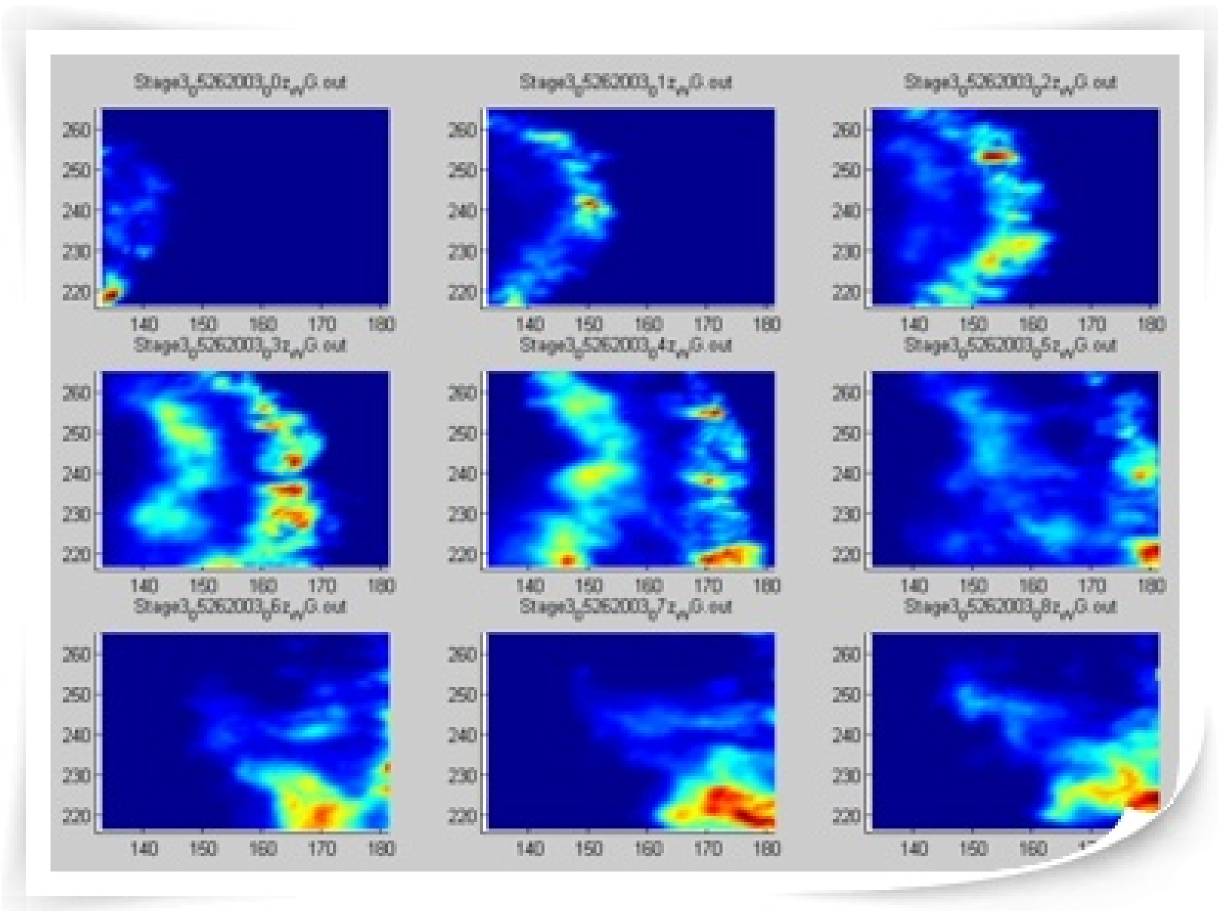


Figure A-5. NEXRAD precipitation images.

2. Genetic Algorithm for determining the optimum ellipse

Basically, the genetic algorithm finds out the ellipse that contains the highest volume of precipitation inside it. The procedure to do this is briefly explained in the following set of figures. Detailed explanation about the mathematical and computational concept to realize this algorithm is provided in a separate document.

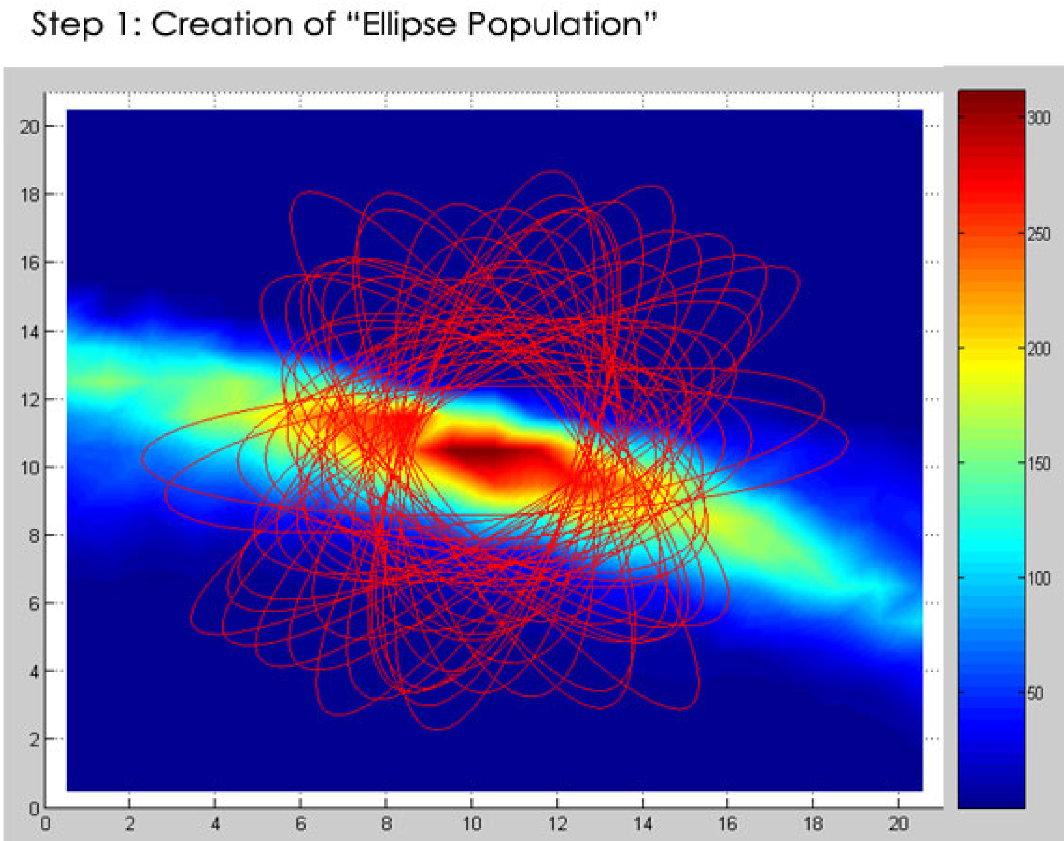


Figure A-6. Creation of ellipse population.

- Changing the value of 1. the inclination angle and 2. the length of the major and minor axis of the ellipse, and generate the "ellipse populations"
- Calculate the average precipitation inside each ellipse (total precipitation inside the ellipse/ellipse area)
- The ellipse with higher average precipitation (higher "fitness") is more likely to be selected as parents based on which next generation is created.

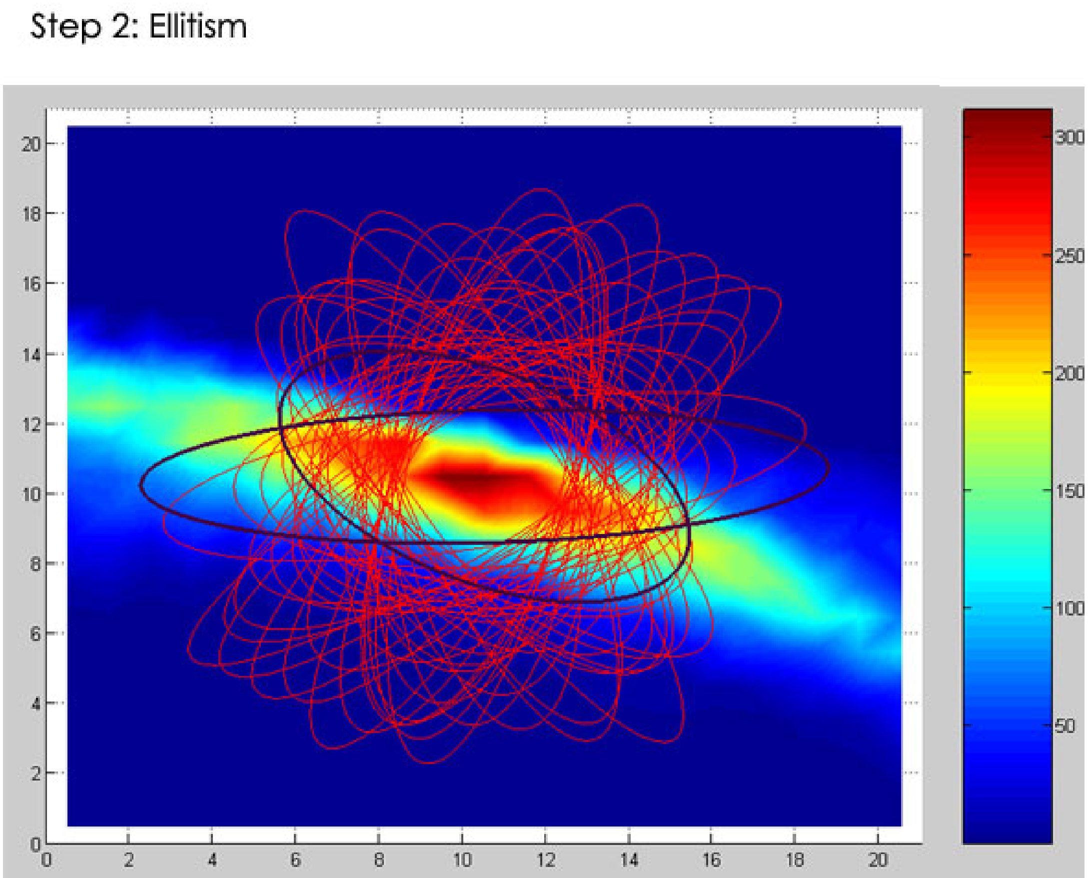
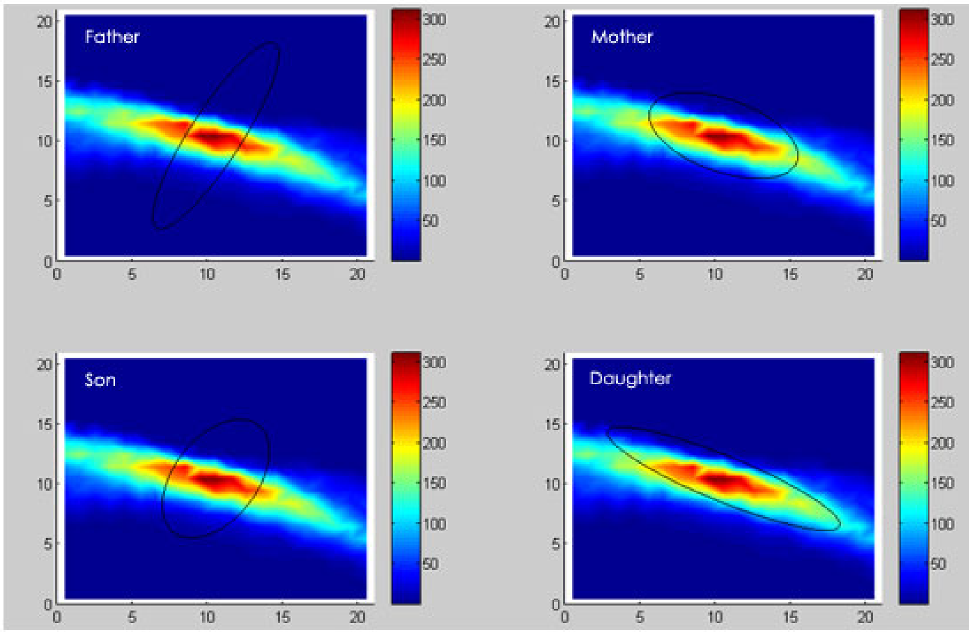


Figure A-7. Elitism.

- Ellipse with the highest fitness (highest average precipitation) is denoted as thick dark blue line.
- These two ellipses are selected as the population of next generation without any genetic modification.
- The process is referred to as “elitism”

Step 3. Crossover and Mutation



Crossover
1. Note that there are two properties that determines the shape thus the fitness of the ellipse;
   a. inclination angle of the ellipse; b. lenght of the major axis. (length of the minor axis is not a variable since the area of the ellipse is fixed). We call these two properties of the ellipse as "chromosome"
2. Randomly pick two ellipses out of initial ellipse population such that the probability of pick is proportional to the fitness of the ellipse. Those ellipses are denoted as "father" and "mother" in the above figure.
3. Interchange the chromosomes of father and mother ellipse to acquire the offspring ellipse - denoted as "son" and "daughter" in the above figure.
4. Son ellipse resembles the inclination angle of the father and major axis of mother, while daugther ellipse resembles the inclination angle of the mother and major axis of father.

Mutation
1. To prevent for a system to converge a local maxima, the chromosomes of the offspring ellipses are sometimes randomly generated regardless of what chromosomes the parents ellipse might have.

Figure A-8. Crossover and mutation.

A.1.3. MATLAB Code Implementation

1. Data Extraction

MPE (Multi-Sensor Precipitation Estimator) data for the year 2003 and 2004 is converted into aforementioned format2 and zipped in the following directory

2003 data → ./2003data/mpe/2003_mpe_mtr.rar

2004 data → ./2004data/mpe/2004_mpe_mtr.rar

Unzip these files in any directory first.

2. Running MATLAB Code

- In the same directory where you unzip the precipitation data files, copy all files in a directory named "../2003data/commonfiles" or "../2004data/commonfiles" depending on a year of the data you want to analyze. One can also unzip the precipitation files in these directories. Basically, all files should be in the same directory.

- Open the file named "timeline_2003.dat" or "timeline_2004.dat" using WordPad or Microsoft Excel (Delimiter is space). It should look like as follows;

stage3_01012003_00z_WG.out 1
stage3_01012003_01z_WG.out 2
stage3_01012003_02z_WG.out 3
stage3_01012003_03z_WG.out 4
stage3_01012003_04z_WG.out 5
……
stage3_01012003_19z_WG.out 8756
stage3_01012003_20z_WG.out 8757
stage3_01012003_21z_WG.out 8758
stage3_01012003_22z_WG.out 8759
stage3_01012003_23z_WG.out 8760

Each line of the file represents the name of the hourly precipitation data file. They should be sorted in ascending order according to time which is the part of the name of the file. Name of the file in the first column should match the files that you extracted in step 2.1. For example, if the name of the files you extracted from the step 2.1 are as follows:

mtr_01012005_00z_WG.mtr
mtr_01012005_01z_WG.mtr

mtr_01012005_02z_WG.mtr

…

mtr_12312005_23z_WG.mtr

The “timeline_2003.dat” file should be as follows;

mtr_01012005_00z_WG.mtr 1

mtr_01012005_01z_WG.mtr 2

mtr_01012005_02z_WG.mtr 3

…

mtr_12312005_23z_WG.mtr 8760

- Open the MATLAB m-file starting with arf_*****.m in the “./2003data/commonfiles” or “./2004data/commonfiles” directory. Executing these MATLAB code will calculate the optimum ellipse. The calculation of the ellipse will be repeated for the following variables:

- Point of annual maxima - another text file that contains the information of annual maxima should be prepared to run the code. The specific explanation is given in the separate document.

- Different sizes of ellipse

- A certain line of the code contains the information about the size of the ellipse that one wants program to calculate. Separate document also contains this information

A.1.4. Examining Relationships of Dependant Variable and Explanatory Variables

The ARFs may be related to storm characteristics (storm area, depth, duration, aspect ratio, and orientation angle, season, and region. Multiple regression analysis was used to sort out what were important factors affecting the independent variable. Analysis of variance (ANOVA) was used to test if the main factors (means) of the ARFs were significantly different for different durations while fixing the area. The other factors (season and region) were investigated to see if they can explain the variance of ARFs while fixing the identified variables (area and duration). Analysis of Covariance (ANCOVA) is conducted to test the effects of season and region on the ARFs with a continuous independent variable, called covariate. The storm properties obtained in Section 2 were stored as a SPSS data file (.sav). The stored data consists of storm occurrence time, NEXRAD grid cell numbers (HRAP X and Y), area (the number of NEXRAD cells), duration, angle (orientation angle), aspect (aspect ratio: ratio of major axis length to minor axis length), and depth at the center of storm. The statistical analysis starts with opening the previously saved data set. The "get file" command in SPSS syntax is used to load the data into SPSS. The syntax in the SPSS Editor is as follows.

GET FILE='C:\Research\TXDOT\Statistical Analysis\arf.sav'.

Regression is a statistical technique that estimates a line that best fits the data. Regression analysis creates two types of statistics. One is the descriptive of independent variables and the relationship between each independent variable and the dependent variable. The other is the statistics of regression model. In regression analysis, the relative contribution of each independent variable can be assessed by the selection methods. The regression procedure in SPSS provides five methods to select predictor variables. They are *enter (forced entry)*, *forward selection*, *backward elimination*, *stepwise selection*, and *forced removal*. The enter method is used first to evaluate the relative importance of each independent variable. From the result, the possible important variables were identified. Once the important variables were identified, the stepwise method is used to create the most parsimonious model. From the regression models created by the selection methods, the most important predictors were selected. The syntax in the SPSS Editor for the regression analysis is as follows.

```
REGRESSION
/DESCRIPTIVES MEAN STEDEV CORR SIG N
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA CHANGE ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT arf
 /METHOD=ENTER area duration depth aspect angle
/RESIDUALS DURBIN.
```

The optional parameters used in the regression analysis are described as follows.

- Descriptive statistics – the means, standard deviations, correlation matrix, significance of each correlation, and the sample size
- Statistics – the unstandardized and standardized coefficients and their significance (T-test results). Multiple R, R^2, adjusted R^2 and standard error of the estimate (R), the ANOVA table, the change of R^2 and their significance, and zero-order, partial and part correlation (ZPP).
- METHOD – Specifies the block of variables to be included as Independent Variables (IVs). Enter (forced entry) tells SPSS to use all IVs simultaneously in the regression equation.

The SPSS syntax for the regression analysis produces the statistical summaries and results as follows. The descriptive statistics of each variable is shown in Table A-1.

Correlation is the covariance of standardized variables. The correlation coefficient ranges from -1 to 1. A correlation coefficient of 0 indicates no linear relationship and -1 or 1 indicates a perfect relationship.

Table A-1. Descriptive statistics.

	Mean	Std. Deviation	N
Arf	.6321	.19872	148271
Area	16.75	14.847	148271
Duration	5.75	6.905	148271
Depth	62.1643	35.45768	148271
Aspect	2.5626	1.18143	148271
Angle	81.5055	52.44684	148271

Table A-2. Correlations.

		Arf	Area	Duration	Depth	aspect	Angle
Pearson Correlation	Arf	1.000	-.588	.132	.124	.068	-.038
	Area	-.588	1.000	.000	.000	.064	-.007
	Duration	.132	.000	1.000	.465	.007	.008
	Depth	.124	.000	.465	1.000	-.033	-.028
	Aspect	.068	.064	.007	-.033	1.000	-.066
	Angle	-.038	-.007	.008	-.028	-.066	1.000
Sig. (1-tailed)	Arf	.	.000	.000	.000	.000	.000
	Area	.000	.	.492	.487	.000	.002
	Duration	.000	.492	.	.000	.002	.001
	Depth	.000	.487	.000	.	.000	.000
	Aspect	.000	.000	.002	.000	.	.000
	Angle	.000	.002	.001	.000	.000	.
N	Arf	148271	148271	148271	148271	148271	148271
	Area	148271	148271	148271	148271	148271	148271
	Duration	148271	148271	148271	148271	148271	148271
	Depth	148271	148271	148271	148271	148271	148271
	Aspect	148271	148271	148271	148271	148271	148271
	Angle	148271	148271	148271	148271	148271	148271

In Table A-2, it was shown that the ARF is most closely related to area (correlation coefficient = -0.588), followed by duration (correlation coefficient = 0.132, showing weak relationship), in agreement with the previous research on ARFs. The depth was also weakly related to ARF (correlation coefficient = 0.124). However, aspect ratio (aspect) and orientation angle (angle) may not be important predictors for the ARFs (correlation coefficient = 0.068 for aspect, correlation coefficient = -0.038 for angle).

Table A-3. Variables entered/removed (b).

Model	Variables Entered	Variables Removed	Method
1	area, duration, angle, aspect, depth(a)	.	Enter

All requested variables entered.
Dependent Variable: arf

The "model summary (b)" table shown in Table A-4 tells us how much the variance of the dependant variable can be accounted for by the independent variables. All requested variables were entered simultaneously. R^2 is a measure of how much of the variability in the outcome. R^2 is accounted for by the predictors. For the initial model its value was 0.381, which means that all predictors account for 38.1% of the variation in ARF. We still have an unexplained variation of about 60 %. Thus, this initial model was not enough to explain the variability of the outcome. The change in the amount of variance explained gave rise to an F-ratio of 18231.851, which is significant with a probability less than 0.001. The value of Durbin-Watson test was .113 less than, 3, and it, thus, can be thought that there is no significant serial correlation of residuals. The Durbin-Watson tests whether adjacent residuals are correlated.

Table A-4. Model summary (b).

Mod-el	R	R^2	Adj. R^2	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.617	.381	.381	.15638	.381	18231.9	5	148265	.000	.113

Predictors: (Constant), angle, area, duration, aspect, depth
Dependent Variable: arf

In the “ANOVA (b)” table shown in Table A-5, the value of F-ratio for the regression model is 18231.851, which is very unlikely to have happened by chance ($p < 0.001$). These results can be interpreted as meaning that the initial model significantly improved an ability to predict the outcome variable.

Table A-5. ANOVA (b).

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2229.307	5	445.861	18231.851	.000(a)
	Residual	3625.832	148265	.024		
	Total	5855.139	148270			

Predictors: (Constant), angle, area, duration, aspect, depth
Dependent Variable: arf

As shown in the following table, as area increases, the ARF decreases. From past research, it is known that the relationship between area and ARF is negatively correlated. It is noticeable that the angle has a negative effect on ARF. The unstandardized coefficient B of the angle is -0.00012670 (the value B shown in the table is truncated so that only 4 digits to the right of the decimal point are considered). Although the value B is the slope of the regression line, it can also be interpreted as the change in the outcome associated with a unit change in the predictor. Therefore, if the angle increases by 1 degree, the ARF decreases by 0.00012670. The allowable range of the value B for an angle is ±180 degree. The maximum difference of orientation angles of two different storms is 90 degree. Thus, if the angle increases by the maximum value 90 degree, the ARF decreases by 0.01. Therefore, it was considered that the slope of the angle was too small to be included in the regression model. The depth, aspect, and duration had a slight tendency to go same direction with ARF.

Table A-6. Coefficients (a).

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	Constant	.686	.001		466.226	.000			
	area	-.008	.000	-.595	-290.470	.000	-.588	-.602	-.594
	duration	.003	.000	.093	40.402	.000	.132	.104	.083
	depth	.000	.000	.083	35.853	.000	.124	.093	.073
	aspect	.018	.000	.106	51.619	.000	.068	.133	.105
	angle	-.000	.000	-.033	-16.314	.000	-.038	-.042	-.033

Dependent Variable: arf

Once the important predictors were identified (area, duration, depth, and aspect), the regression analysis was re-run to include only the important predictors and checked if they substantially improve the model's ability to predict the outcome. At this time the regression predictors were selected based on past research and the results of the regression analysis. It was decided that the predictors were entered into the model in a hierarchical way. Two selection steps were used to select the predictors in the models. At the first step, the known predictor "area" was entered. At the next step, the others (duration, depth, and aspect) were entered simultaneously using stepwise selection method. The stepwise method in SPSS searched for the predictor that best predicts outcome variable based on the highest simple correlation between the outcome and the selected predictor. If this predictor significantly improves the ability of the model to predict the outcome, then this predictor is retained in the model. It continues until all predictor variables were tested. The syntax to do this regression analysis is as follows.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE ZPP
```

```
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT arf
/METHOD=ENTER area  /METHOD=STEPWISE duration depth aspect
/RESIDUALS DURBIN .
```

This regression analysis produced 4 different regression models shown in Table A-7. The first model (1) included area as a single predictor variable. The second model (2) added duration as predictor variable so this model has two predictor variables (area and duration). The third model (3) added aspect as predictor variable, while retaining area and duration. Finally the last model (4) added depth as predictor variable, while retaining area, duration, and aspect. All the models were shown to be statistically significant in the F change (all significance levels were less than 0.0001), shown at the column "Sig. F change" in the table. Interpretation of the statistical results, however, should be careful in a sense that the narrow range of ARF (bounded between 0 and 1), and a large sample size lead to small significance level for F changes. It was necessary to check the R square change along with the significance level of F change. The R square changes by introducing the model 2, 3, and 4 were so small (0.018, 0.011 and 0.006 for each) that it was difficult to say that they improved the ability to predict the outcome. Thus, it was concluded that only area is the important predictor in the regression model.

Table A-7. Model summary (e).

Model	R	R^2	Adj. R^2	Std. Error of the Est.	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.588	.346	.346	.161	.346	78273.7	1	148269	.000	
2	.603	.363	.363	.159	.018	4079.9	1	148268	.000	
3	.612	.374	.374	.157	.011	2610.4	1	148267	.000	
4	.616	.380	.380	.157	.006	1331.5	1	148266	.000	.109

A.1.5. Analysis of Variance

It was found that of all of the predictors the area correlated best with the ARFs and so it is likely that the area will best predict the ARFs. The other predictors except area, however, were very weakly related to the ARFs. The adjusted R^2 of the model 1 is 0.346 and it tells us that 35% of the overall variance is explained by the model. The unexplained variance, however, needs to be explained by taking into account other factors such as season and region. This dissertation is focusing on the assumption that the other factors can be used to explain the variance of ARFs, rather than focusing on building the regression model. The duration can be considered as a categorical variable, and then Analysis of Variance (ANOVA) test can be used to test if the means of ARFs at a specified level of area were different for different durations (6 groups or levels). Thus, the duration was further investigated to see if the variance of ARFs can be explained by the duration while fixing the area at a level. The area consists of 8 different levels (3, 5, 7, 9, 11, 19, 31, and 49). Additionally, it was investigated if there were seasonal and regional effects on the variance of the ARFs while fixing the area and duration.

In order to see the effects of duration on ARFs, the area was fixed at the level 7 (area =49). The following scripts are executed in the SPSS editor.

```
USE ALL.
COMPUTE filter_$=(area =49).
VARIABLE LABEL filter_$ 'area =49 (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

UNIANOVA
  arf  BY duration
  /METHOD = SSTYPE(3)
```

/INTERCEPT = INCLUDE
/POSTHOC = duration (BTUKEY BONFERRONI T2)
/EMMEANS = TABLES(OVERALL)
/PRINT = DESCRIPTIVE ETASQ OPOWER PARAMETER HOMOGENEITY
/CRITERIA = ALPHA(.01)
/DESIGN = duration .

When ANOVA test is applied, the assumption that the variances in each of the groups are same should be met. The Levene's Test tests equality of error variances of the dependant variable across groups. Based on p-value (0.001) less than the significance level (α = 0.01) in Table A-8, it failed to reject the null hypothesis that error variances were equal across groups. Even though equal variances are not valid in this model, the results using one-way ANOVA test are approximately correct if the sample sizes are large. An alternative test (e.g., Wilcoxon ran sum test) that requires less strict conditions can be used instead.

Table A-8. Levene's test of equality of error variances(a).
Dependent Variable: arf

F	df1	Df2	Sig.
23.300	4	18526	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

One way ANOVA statistics are shown in Table A-9. Duration is a factor which is an independent variable. The analysis of variance to analyze the effect of duration is called one-way or one-factor ANOVA. The corrected model is the overall model. It includes the variance due to all of the effects in the design.

The effects consist of main effects and interaction effects. The main effect is the simple effect of a factor on a dependant variable. The main effect is obtained by averaging the factor across the levels of other factors. The interaction is the effect of a factor depending on a level of the other factor. One way ANOVA only have main effects so that the interaction terms are not included in the model.

The "Sig." column in Table A-9 gives the probability (p) value of the F test. Since the p value is less than the significance level ($\alpha = 0.01$), it is concluded that duration does make a difference in ARFs.

Table A-9. Tests of between-subjects effects.

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power(a)
Corrected Model	53.084	4	13.271	458.490	.000	.090	1833.962	1.000
Intercept	2951.46	1	2951.46	101967.98	.000	.846	101967.976	1.000
Duration	53.084	4	13.27	458.490	.000	.090	1833.962	1.000
Error	536.23	18526	.029					
Total	3940.33	18531						
Corrected Total	589.32	18530						

Computed using alpha = .001
R Squared = .090 (Adjusted R Squared = .090)

Post-hoc tests are used to assess which group means differ from which others. All possible pairwise comparisons between means were obtained by one-way ANOVA test in SPSS. The Bonferroni test and Turkey HSD test results were shown In Table A-10. The mean of each group was compared to the mean of the other. The null hypothesis for this test is that there is no difference in the means of two groups. If the probability value (p-value) is less than the significance level (α =0.05), the null hypothesis is rejected. At the first row, the mean difference (I-J) of duration 1 and 3 is -0.0716(*). The asterisk indicates that the mean difference is significant at the 0.01 level. This means that the null hypothesis is rejected so that there is difference in the means of two groups. All cases were found to be significant in mean difference. It was concluded that the duration could reduce the variance of ARFs.

Table A-10. Multiple comparisons.
Dependent Variable: arf

	(I) duration	(J) duration	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interval	
						Lower Bound	Upper Bound
Bonferroni	1	3	-.0716(*)	.00328	.000	-.0824	-.0608
		6	-.1018(*)	.00372	.000	-.1140	-.0895
		12	-.1230(*)	.00415	.000	-.1366	-.1093
		24	-.1415(*)	.00447	.000	-.1562	-.1268
	3	1	.0716(*)	.00328	.000	.0608	.0824
		6	-.0301(*)	.00411	.000	-.0436	-.0166
		12	-.0513(*)	.00450	.000	-.0661	-.0365
		24	-.0699(*)	.00480	.000	-.0857	-.0541
	6	1	.1018(*)	.00372	.000	.0895	.1140
		3	.0301(*)	.00411	.000	.0166	.0436
		12	-.0212(*)	.00483	.000	-.0371	-.0053
		24	-.0398(*)	.00511	.000	-.0566	-.0230
	12	1	.1230(*)	.00415	.000	.1093	.1366
		3	.0513(*)	.00450	.000	.0365	.0661
		6	.0212(*)	.00483	.000	.0053	.0371
		24	-.0186(*)	.00543	.006	-.0364	-.0007
	24	1	.1415(*)	.00447	.000	.1268	.1562
		3	.0699(*)	.00480	.000	.0541	.0857
		6	.0398(*)	.00511	.000	.0230	.0566
		12	.0186(*)	.00543	.006	.0007	.0364

Based on observed means.
* The mean difference is significant at the .01 level.

SPSS produces a table which lists all the groups and their means. For the significance level of 0.01, additional columns are added. Each column is the one for each subset where group means do not differ significantly. If the probability value (p-value) is less than the significance level (α-value), we fail to accept the null hypothesis that there is no difference in the means of groups. The range of tests in Table A-11 identified subsets of means that differ from each other. There were 5 subsets for the ARFs and all durations were significantly different from each other.

Table A-11. Homogenous subsets.

	duration	N	Subset				
			1	2	3	4	5
Tukey B(a,b,c)	1	7479	.3650				
	3	4189		.4367			
	6	2895			.4668		
	12	2173				.4880	
	24	1795					.5065

Means for groups in homogeneous subsets are displayed. Based on Type III Sum of Squares The error term is Mean Square(Error) = .029.
Uses Harmonic Mean Sample Size = 2881.599.
The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
Alpha = .01.

A.1.6. Seasonal and Regional Effects on ARFs

As with the previous research on ARFs, ARFs was closely related to area. It was also found that the duration can explain the variance of ARFs for each level of area. In order to further reduce the variance, the season and region along with depth were taken into account.

1. Analysis of Covariance (ANCOVA) Test

In order to investigate the effects of season and region on the variation of ARFs, ANCOVA was used to test the main and interaction effects of categorical variables (season and region) on a continuous variable (ARFs) while controlling for the effects of the other continuous variable (depth) which covariates with the dependant (ARFs), also called the covariate. The statistical tests proceeded in two steps for season. First, for each season, ANCOVA was used to test the effects of season on ARFs with depth. Second, homogeneous subsets obtained from the multiple comparisons were used to group the seasons that are not significantly different each other. Likewise, ANCOVA was used to test the effects of region on ARFs with depth while fixing the season. In SPSS, the ANCOVA is part of ANOVA procedure. Univariate General Linear Model

(GLM) is the version of the general linear model often used to implement ANOVA and ANCOVA.

2. Seasonal Effects on ARFs

After subdividing the dataset into 40 subsets of records of the same duration and area (i.e., 5 × 8 combinations), sinusoidal regression equations of the ARF with the storm time of occurrence as predictor were developed. The sinusoidal equations had the form

$$ARF = a_1 - a_2 \sin\left(2\pi \frac{H - a_3}{H_{Max}}\right)$$

A SPSS Visual Basic script file (SAToolD.SBS) is created to calculate the sinusoidal equations. The sub procedure for the derivation of the sinusoidal equations is $main_sinusoidal. All sub procedures in SPSS VB script file (.SBS) must be executed inside the sub procedure $main. For example, in order to run $main_sinusoidal the scripts are written shown in Figure A-9 in the Visual Basic script editor window.

To run a script, open the SPSS scripting facility from the File menu.

File - New - Script

The sinusoidal phase shifts were read from the pivot tables of the SPSS output (.spo) of the $main_sinusoidal, and written into a text file by the function $main_sinusoidal_phaseshift to obtain the mean of the phase shifts. The output text file was stored in the directory ./text. The mean was used to divide the year into two six-month seasons: summer and winter.

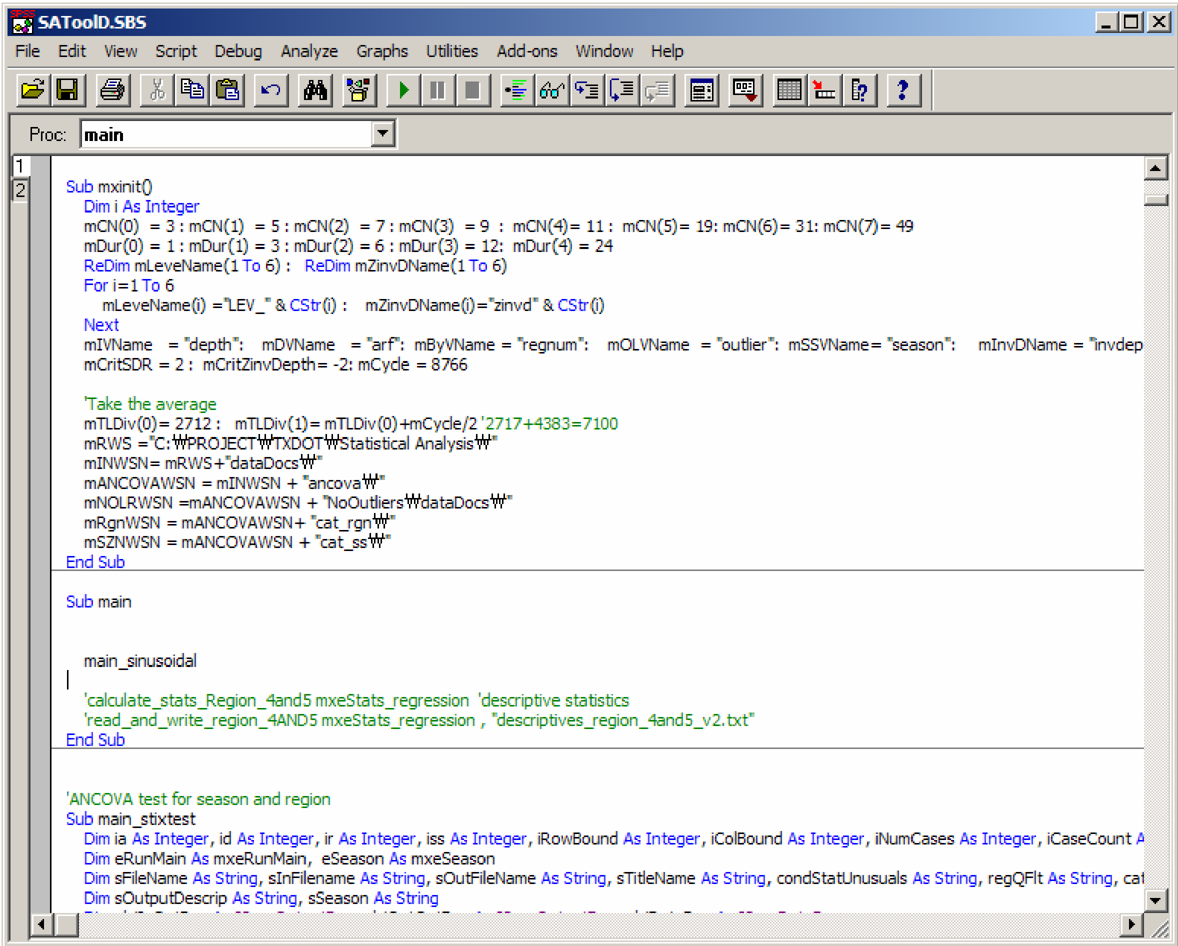


Figure A-9. Run main_sinusoidal (the script in the red box).

To reduce processing time for the statistical analysis, the 40 subsets of records of the same duration and area were stored in SPSS .sav files (compatible with SPSS 6.0 and later). The file name was specified as a<area>d<duration>.sav where <area> is the numeric value for area (the number of NEXRAD cells: 3, 5, 7, 9, 11, 19, 31, and 49), <duration> is the numeric value for duration (1, 3, 6, 12, and 24 hours). Example file names are as follows:

a3d1.sav – area (the number of cells) = 3 and duration = 1

a31d12.sav – area (the number of cells) = 31 and duration = 12

ANCOVA test for the season as a factor and the depth as a covariate was conducted by the function $main_stixtest::$main_ANCOVA_season_depth. The output file was saved as d(#1)oc(#2)csu(#3)r(#4)ss(#5).spo where #1 is the duration value, #2 is the number of the outlier criteria used, #3 is either "or" or "and" (used when defining and selecting outliers by the conditional operations among the outlier criteria), #4 is the region number, and #5 is either 0 (summer) or 1 (winter) or 2 (both included in the dataset). The listed output is only for the case of duration 1, region 1, and all areas. SPSS provides several diagnostic statistics that allow the case-by-case evaluation of the data for possible influential cases. The outlier diagnostics used were Studentized Deleted Residual, Cook's distance, Standardized DFFIT, and Standardized DFBETA. The description about the outlier diagnostics is as follows.

- **Cook's distance**. Measure how much the estimated regression coefficients would change when *i*th observation is deleted. As rule of thumb, any observation with the distance greater than 1 has high influence.
- **Studentized Deleted Residuals**. Residuals divided by the standard error of the residual with that case deleted. Regression equation is recalculated with the set of data excluding the observation in question. As a rule of thumb, studentized deleted residuals > 2 is uncommon and those > 3 is rare, considered to be outlier in this study.
- **Standardized DFBETA.** Standardized difference in beta value. The change in the regression coefficient that results from the exclusion of a particular case. It is considered outlier in case which its absolute value of standardized difference is greater than 2 divided by the square root of N, where N is the number of cases. A value is computed for each term in the model, including the constant.

- **Standardized DFFIT**. Standardized difference in fit value. The change in the predicted value that results from the exclusion of a particular case. It is considered outlier in case which its standardized value in absolute value exceeds 2 times the square root of p/N, where p is the number of parameters in the model and N is the number of cases.

The $main_stixtest::$main_ANCOVA_season_depth is an interaction model and it represents the interaction effect by adding an additional term (interaction term) in the model. If the interaction is not significant which means the coefficient of interaction term goes to zero, it is necessary to re-run ANCOVA test without interaction effects to get the correct set of parameters. The results of ANCOVA test were represented as three different types: existence of difference in means, non-existence of difference in means, and existence of interaction. The type was represented as an index (0, 1, and 2) in the same sequence. The function $main_stixtest::main_multcmp created a text file that contains the type as index. The output was stored as rtype.txt in the directory /ancova/cat_ss/reports. The coefficients of the regression equations were obtained by the function $main_stixtest::$main_calcregparams based on the types obtained from the rtype.txt. The outputs of the function were saved as coeffs calc-d(#1)oc(#2)r(#4)ss(#5).spo in the directory ./ancova/cat_ss/outputDocs. There were 35 SPSS output files created by the function. The listed output is only for the case of duration 1, region 1, and all areas (i.e., coeffs calc-d(1)oc(2)r(1)ss(2).spo).

3. Regional Effects on ARFs

Texas was divided into six regions based on the climate conditions. The storm events belonging to the each region are assigned a code ranging from 1 to 6 for the statistical analysis of the groups. The statistical tests for region follow the same procedure as for the seasons. The function $main_ANCOVA_region_depth was used to test the main effects of region on the

ARFs with depth. ANCOVA test for the region as a factor and the depth as a covariate was conducted by the function $main_stixtest::main_ANCOVA_region_depth. The output file was saved in the directory ./ancova/cat_rgn/outputDocs as d(#1)oc(#2)csu(#3)ss(#4).spo where #1 is the duration value, #2 is the number of the outlier criteria used, #3 is either "or" or "and" (used when defining and selecting outliers by the conditional operations among the outlier criteria), and #4 is either 0 (summer) or 1 (winter) or 2 (both included in the dataset). The listed output is only for the case of duration 1, summer, all regions, and all areas. The function $main_stixtest::$main_multcmp created a text file that contains the type as index. The output was stored as rtype.txt in the directory /ancova/cat_rgn/reports. The coefficients of the regression equations were obtained by the function $main_stixtest::$main_calcregparams based on the types. The outputs of the function were saved as coeffs calc-d(#1)oc(#2)r(#4)ss(#5).spo in the directory /ancova/cat_rgn/outputDocs. The outputs of the function are saved as coeffs calc-d(#1)oc(#2)r(#4)ss(#5).spo in the directory /ancova/cat_rgn/outputDocs.

A.2. THE DISAGGREGATION MODEL

The programs used for the disaggregation model are FORTRAN, SPSS 11.0.1, MATLAB 7.0, and ArcGIS 9.0.

A.2.1. Directory Structure

All the input, output and script files for the analysis are stored in the top-level directory and its subdirectories. There are two top-level directories. One stores the inputs, outputs, source codes, and associated files of MATLAB. The other stores the inputs, outputs, source codes, and associated files of FORTRAN, SPSS, and ArcGIS program. The directories are listed below.

1. The top-level directories for FORTRAN, SPSS, and ArcGIS programs (c:/research/raindis/)

./	ArcMap project files (.mxd), Geodatabases (.mdb)
./statistics	Contain subdirectories, and in each subdirectory there are SPSS data files (.sav), output files (.spo), SPSS scripts (.SBS), and text files (.txt) for the statistical analysis
./shape	Shape files (.shp, sbx, sbn, prj and dbf)
./geostatistics	Contain subdirectories and the results of the Geostatistical Analyst
./documents	Documents for report (.doc, .ppt, and .xls)
./arcScripts	ArcGIS Visual Basic Script files (.bas, .frm, and .cls)
./data	Contain subdirectories and data files (.txt, .dat, .dbf, and .xls)
./FORTRN	Contain subdirectories, FORTRAN programs and the associated files

Subdirectories

The section provides information about some of the subdirectories and their contents.

./FORTRN

./bin	executable files
./Data	raw data
./Results	Subdirectories and outputs of FORTRAN program
./src	Source codes of FORTRAN program and C program (.f and .c)

./FORTRN/Results

./	Directories named as tx000 for gauge 000
./tx001/	Directory for gauge 001
./tx/	Texas state database

./statistics

./dataDocs	SPSS data files (.sav)
./outputDocs	SPSS output files (.spo)
./scripts	SPSS script files (.SBS)
./text	Text files (.txt), containing the outliers of epsilon values for each month

./geostatistics

./	Directories named as eps00 for month 00
./eps1	Directory for month 1

2. The top-level directories for MATLAB (c:\MATLAB7\work\toolbox\)

../	Common MATLAB m-files (.m) for the three methods (the derivation of ARFs, the disaggregation method, and the storm feature identification and tracking method)
./	MATLAB m-files (.m)
./statlearn	MATLAB objects and functions modeling multivariate densities and classification procedures
./stix	functions for statistical analysis
./storm	Contain subdirectories are related to the methods presented in this dissertation, and common functions (.m) shared by the methods

Subdirectories

The section provides information about some of the subdirectories and their contents.

./statlearn

./unsupervised	MATLAB objects (density, gaussdens, diagdens, balldens, pca-dens, mixdens, unifdens, and mxmoddens)

./storm

./dsg	Contain subdirectories and functions associated with the rainfall disaggregation

./sfit	Contain subdirectories and functions associated with the storm feature identification and tracking algorithm

./storm/dsg

./	MATLAB script files (.m) and MATLAB figure files (.eps and .fig)
./data	MATLAB data files (.mat) for observed statistics and error statistics

./storm/sfit

./	MATLAB script files (.m) and MATLAB figure files (.eps and .fig)
./data	MATLAB data files (.mat) for storm feature objects and storm feature paths

A.2.2. General Procedure for the Disaggregation Model

The flow chart of Figure A-10 illustrates the general procedure for the disaggregation model. The programs were written in FORTRAN, ArcGIS VBA, and MATLAB. The programs were broken down into six main steps that correspond to the blocks in the flow chart. The flow chart, shown in Figure A-11, illustrates detailed procedures, followed by the flow chart of main steps.

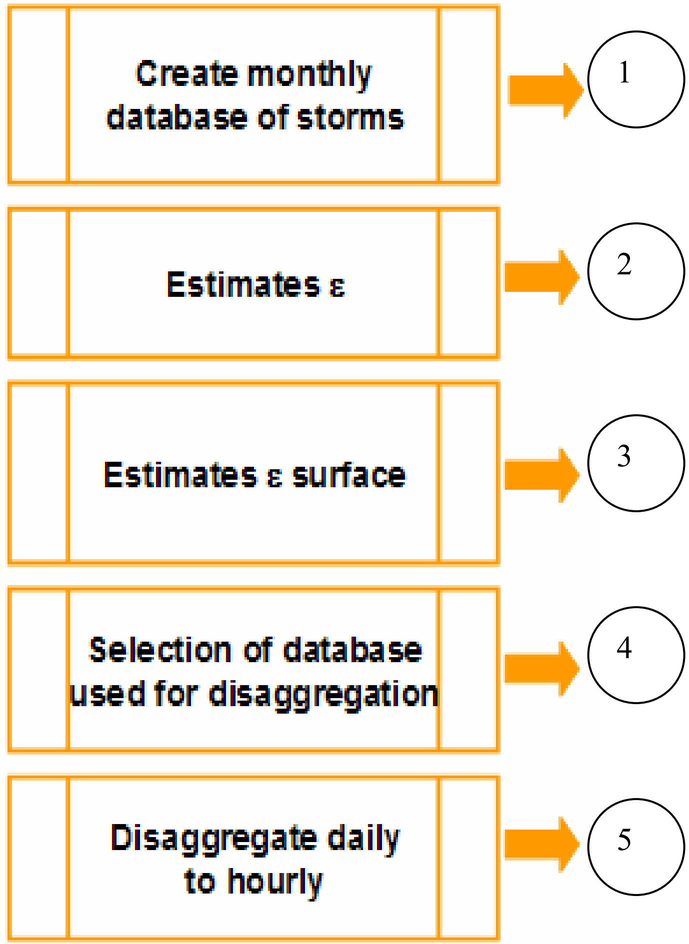


Figure A-10. Flow chart for general procedure of the disaggregation model.

①. Storm is defined as consecutive non-zero rainfall.

②. ε is physically related to the smallest expected one-hour storm event, and is obtained by a calibration.

③. ε is an interpolated estimate.

④. Find the spatial patterns from database.

⑤. Disaggregate daily to hourly rainfall based on best estimated ε and the selected database.

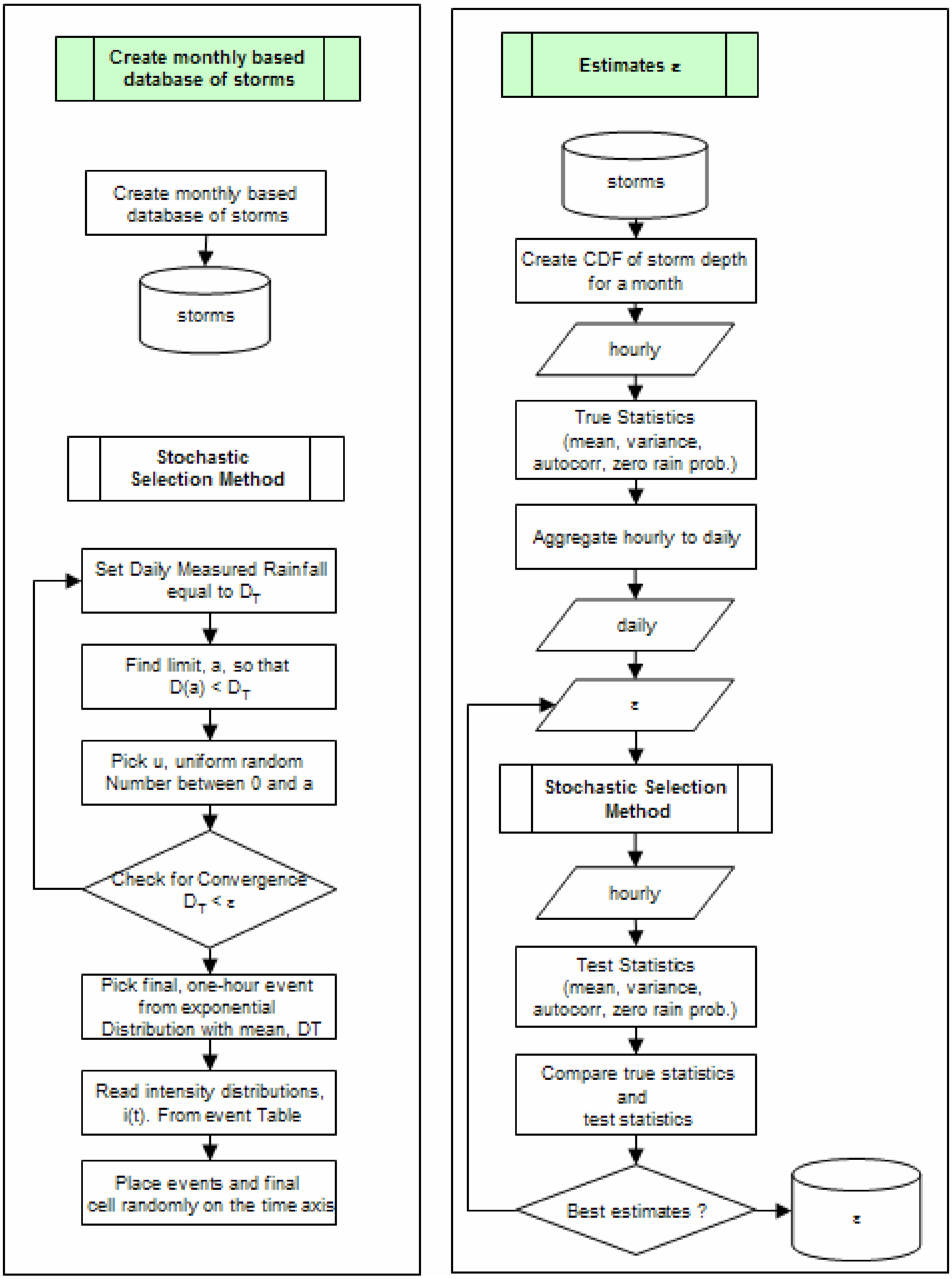


Figure A-11. Flow charts for detailed procedure of the disaggregation model.

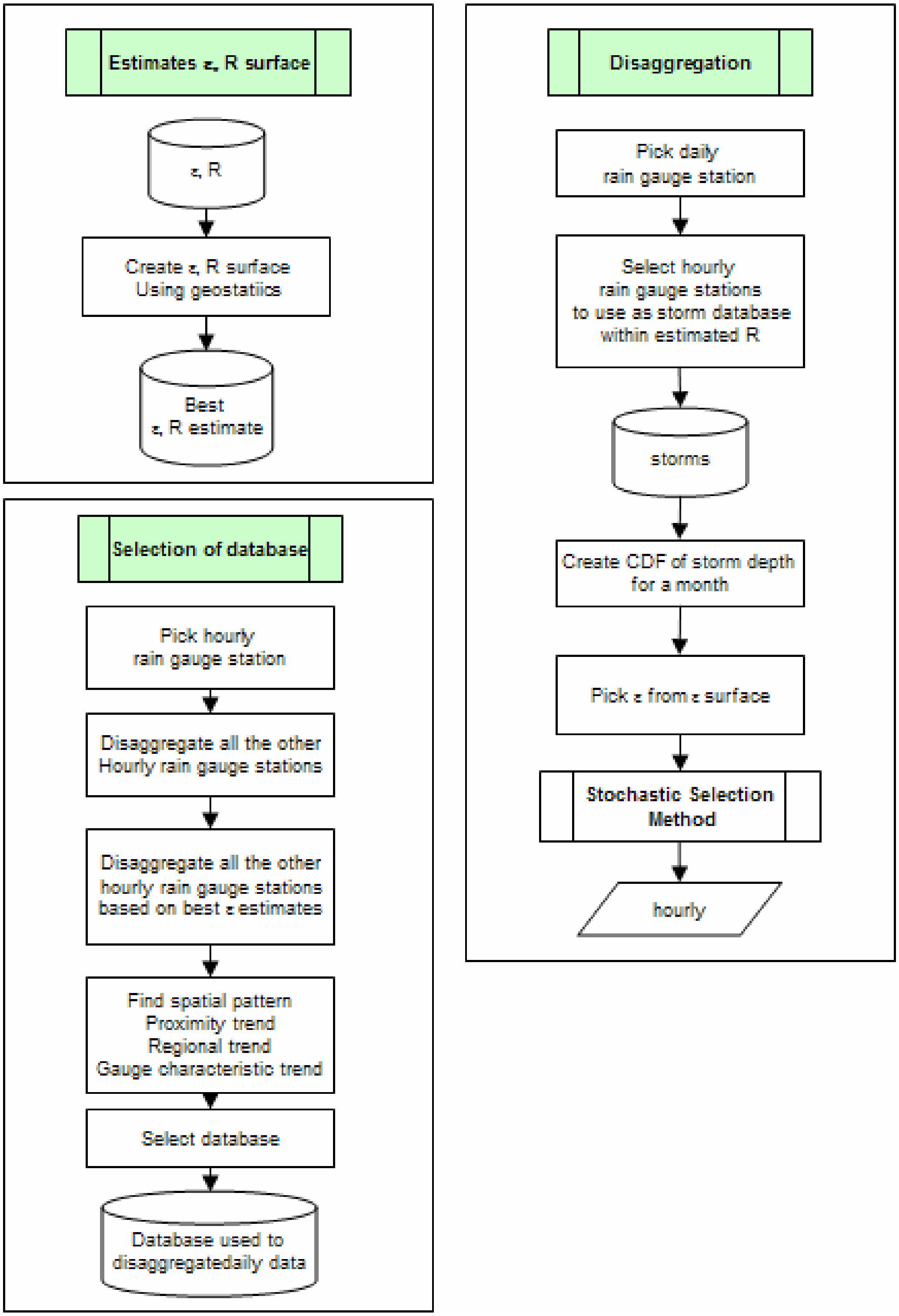


Figure A-11. Continued.

In the first block, a database of storm events (defined as sequences of uninterrupted hourly rainfall) was constructed from the hourly recorded data. Each of the storms for a given month was ranked by their total volume to obtain the cumulative density function for storm volume for the month. The graphical explanation is shown in Figure A-12.

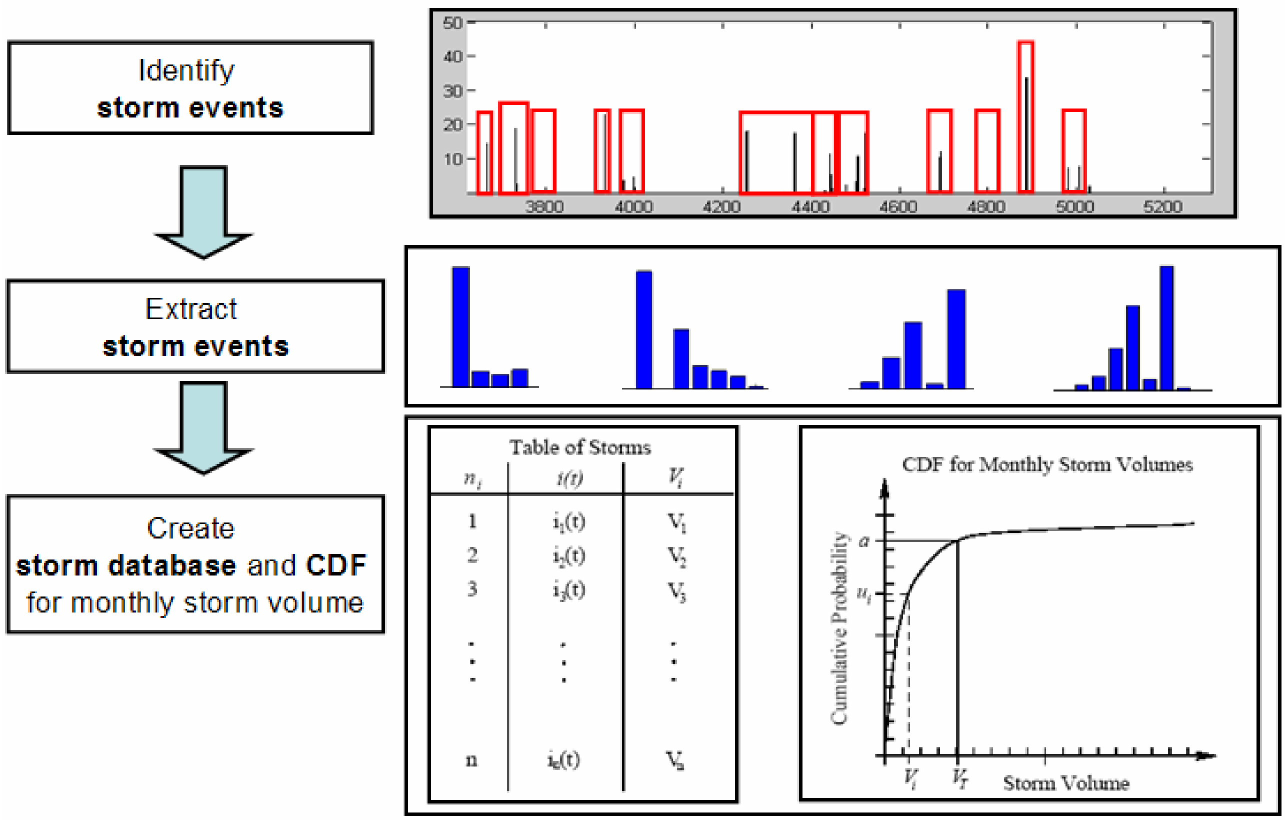


Figure A-12. Construction of hourly storm database.

In the second block, an automated calibration method was applied to obtain the threshold storm size parameter ε. For each month, the smallest-storm parameter ε for each of the 531 gauges was estimated such that it minimized the squared difference between the statistics of the actual precipitation time series (i.e. measured statistics) and the statistics of the time series obtained by auto-disaggregation (i.e. auto-disaggregation statistics). The graphical explanation is shown in Figure A-13.

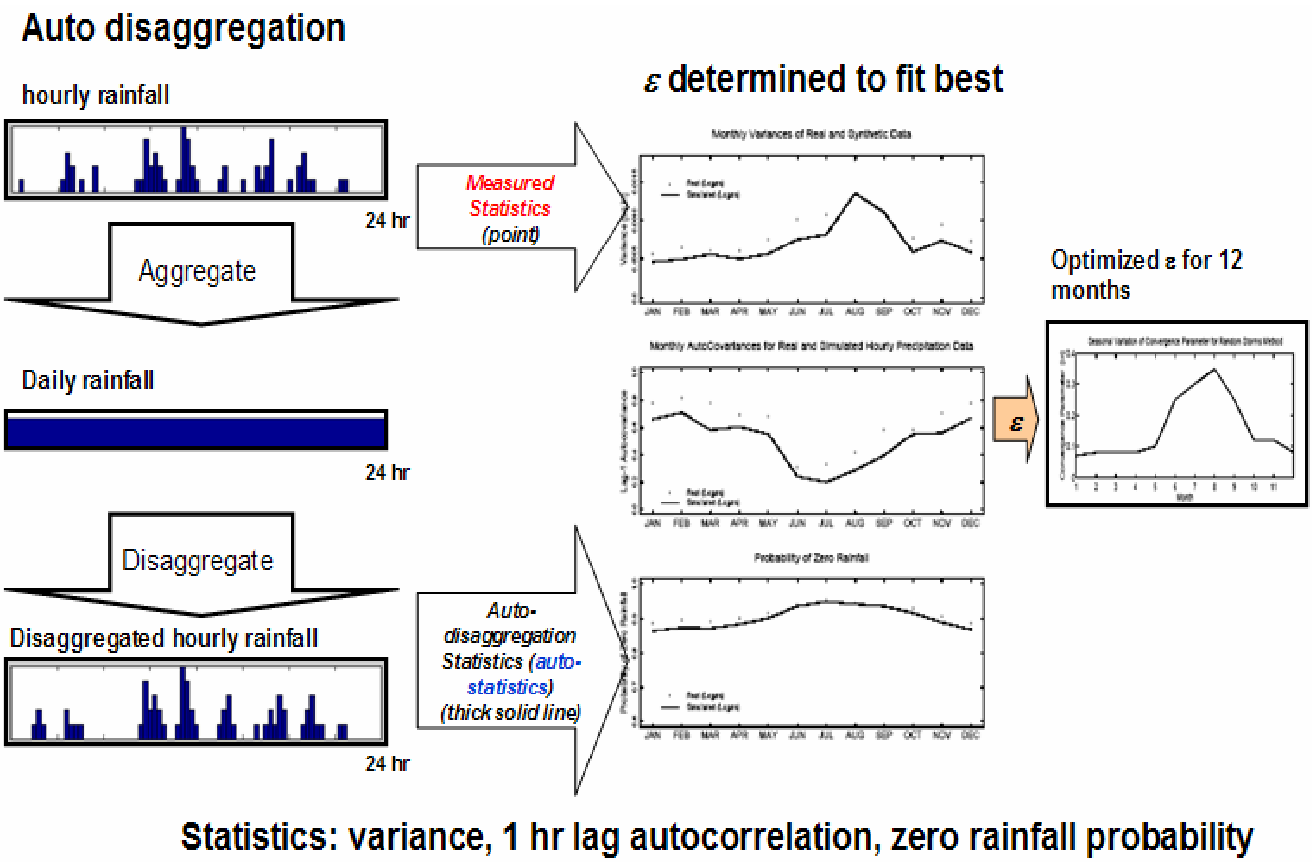


Figure A-13. Estimate best ε for each month.

In the third block, once the ε values were estimated for each month and gauge point, 12 interpolated surfaces were developed, so that ε values could be estimated for each month and at every location in Texas. The interpolated ε surfaces were obtained by Geostatistical Analyst which is an ArcGIS extension. The detailed description of the application procedure is described in the later section.

In the fourth block, three different types of gauge performance trends were investigated for the selection of gauge stations to be used as hourly storm database for disaggregation.

In the last block, once the best estimate of ε and hourly storm database were identified for a gauge station to be disaggregated, the disaggregation method can be applied. The graphical explanation is shown in Figure A-14.

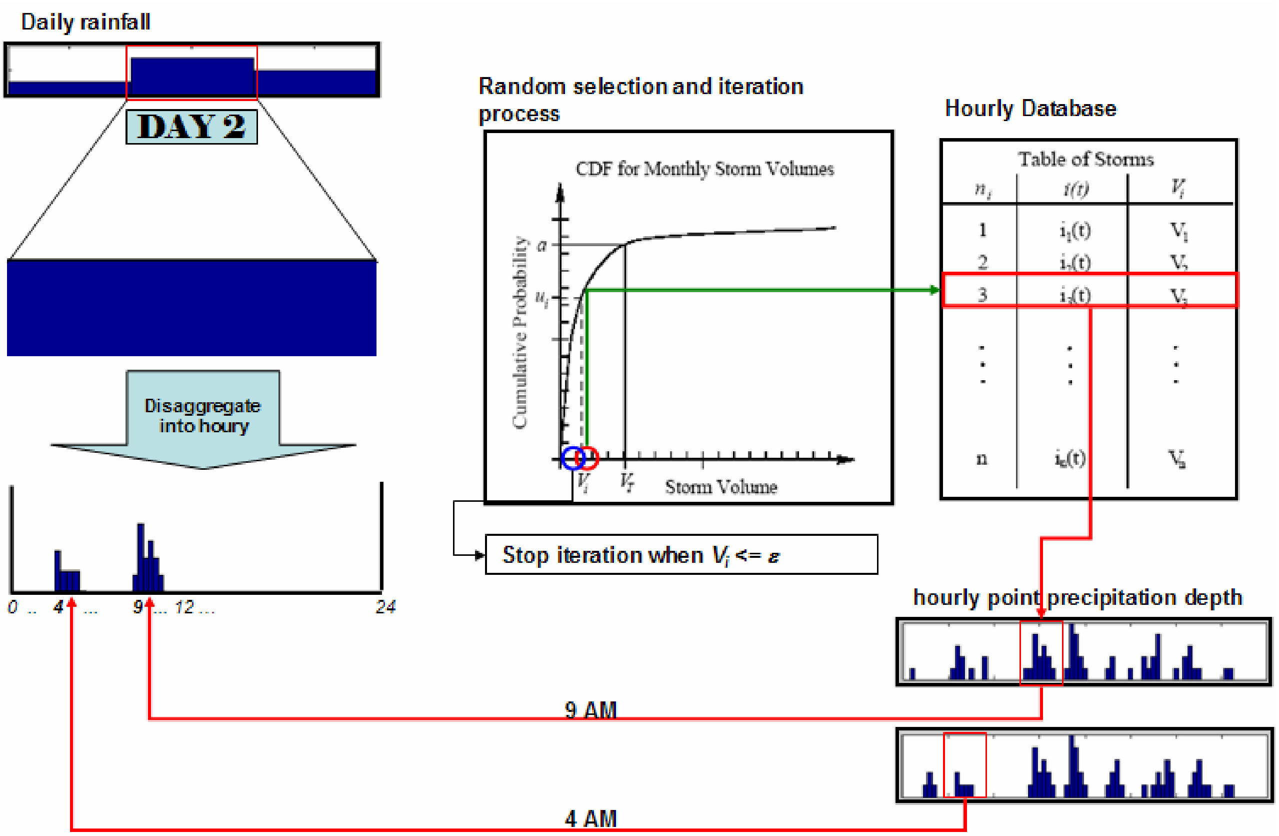


Figure A-14. Diagram for showing disaggregation of daily to hourly.

A.2.3. Running the Programs and its Outputs for the Disaggregation Model

In this section A.2.3 the programs to be executed and their outputs are listed by number in a sequential order. Also, the directories that store the outputs are briefly described. There are some programs to be covered more specifically. Thus, more detailed explanations for the programs are followed by the later section.

1. Remove all the data from the ./results directory

2. Run **mkdirs.bat** in the DOS prompt to create the directory “results” and its subdirectories under the ./FORTRN in which the data will be stored

3. Run #**hp_convert** ../data/tx_rainfall.dat in the DOS prompt to read in the raw data. Two output files will be created:

./results/tx.dat contains the description and location of the stations

./results/tx000/tx000.dat contains the hourly rainfall data for gauge 000

4. Run #**hp_stats** in the DOS prompt to compute the statistics of the hourly data.

One output file (./results/tx000/tx000.stx) which contains the monthly average statistics is created

5. Run #**hp_events** in the DOS prompt to create the databases of storm events.

Monthly output files with the extension .sdb are created in each directory ./results/tx000/.

6. Run #**hp2daily** in the DOS prompt to create daily data from the hourly data at each gauge.

The new daily data file is stored in ./results/tx000/tx000.dly

7. Run #**hp_disagg_baysian** in the DOS prompt to create the optimized values of epsilon at each gauge using auto-disaggregation.

8. Create the epsilon surface in GIS and create the interpolated output file.

This part is explained in detail in the section A.2.5.

9. Then run #**eps_convert** ../data/eps_file.dat in DOS prompt

For example, **eps_convert ../data/rzgaepsilon.dat** that converts the GIS output file to a format readable by FORTRAN

10. Run #**hp_disagg** in the DOS prompt to disaggregate each gauge using the epsilon values interpolated from the epsilon surface created by the Geostatistical Analyst in ArcMap.

Output from #**hp_disagg** is summarized as follows:

In the directories ./results/tx00a/ the following files are created (note that gauge ‘a’ is the daily gauge and gauge ‘b’ is the storm database.):

- TX00a.dat: contains all of the hourly rainfall data. The first three columns are month, day, year, followed by 24 columns of hourly rainfall.
- TX00a.stx: contains the statistics of the measured hourly rainfall. There are 12 columns (January through December). The rows are p0, avg, var, and autocor.
- TX00a_01.sdb through TX00a_12.sdb: store the storm databases for each month (January = 1 through December = 12). The data are sorted by volume and each column is: the total storm depth, number of rainy hours, start time, stop time, and hourly depth for each rainy hour
- TX00a.dly: stores the daily rainfall in month, day, year, rainfall format
- TX00a_00a.sim: stores the simulated data from the auto-disaggregation using the Baysian optimization scheme. The file is in the same format as TX00a.dat.
- TX00a_00a.stx: stores the statistics of the auto-disaggregated simulation. It has the same format as TX00a.stx except that the first row contains the optimized value of epsilon.
- TX00a_00b.stx: stores the statistics of the disaggregated values using gauge b to disaggregate gauge a. They are in the same format as TX00a_00a.stx.

In the directory ./results/ the following matrices of complete results are stored:

- tx.dat: this file contains columns of: user-defined gauge number, NOAA gauge number, start time in month day year, stop time in month day year, and latitude and longitude in degrees minutes seconds
- epsilon.dat: contains the gauge number and 12 columns of epsilon values. This file is created by #**eps_convert**.
- opt_eps.out: contains the optimized values of epsilon calculated through the auto-disaggregation. The first two columns are the gauge number and the storms database number. The remaining 12 columns are the monthly values of epsilon (January through December).
- stats.dat: contains rows for each simulation. The columns are daily gauge, disaggregating gauge, and 12 columns each for epsilon, p0, avg, var, and autocor.

11 .Run #**summary** in the DOS prompt to read each of the output files from step 9 and compile them into a single summary file for use in GIS. The #**summary** create ./results/stats.dat

12. Run @mRainDisaggregation.CreateStatsDifferenceOutputFile to break the stats.dat file into 531 text files. The function is stored in ArcMap mxd file called rgs.mxd. This function reads ./results/stats.dat and creates 531 text files in which the 4 statistics of disaggregated hourly rainfall data are contained

A.2.4. Texas State Database

The results of the disaggregation using the Texas state database are stored in ./results/tx/ in the following files:

- ver_statsions.dat

  This file contains each of the gauges excluded from the Texas Storm database which are to be used for disaggregation. The gauge numbers correspond to user-defined original gauge numbering system 1 through 646. No gauges with less than 5 years of data or missing location information were eligible to be a verification gauge.

- tx.dat

  Raw rainfall data used to create the statewide storm database.

- tx_00.sdb

  These twelve files contain the storms data bases for each month for the state of Texas storm database. This database is used to disaggregate all the other gauges.

- tx000.stx

  These files summarize the fit statistics for the disaggregation of gauge 000 using the state of Texas storm database. These data are combined together in the stats.dat file.

- stats.dat

  This file is organized like the previous stats.dat file. The first column is the gauge number of the daily rainfall gauge. The second column is just zero, indicating that the state rainfall database was used for disaggregation. The remaining columns are twelve each for p0, ave, var, and autocor.

A.2.5. Creation of Epsilon Surface in GIS and the Interpolated Output File

1. Add X, Y coordinates in ArcMap

   Described in detailed in the section A.2.6

2. Resize optimized epsilon values

   1) Open ArcMap file rgs.mxd and run the function @mRainDisaggregation.write_rzeps. This function reads the opt_eps.out file and creates rzopt_eps.out.

3. Identify the outliers of epsilons

   1) Open SPSS and load rzopt_eps.out.

   2) Save it as rzopt_eps.sav

   3) Open SATool.SBS in the ./scripts and run $main_identifyEpsilonOutliers

   The $main_identifyEpsilonOutliers creates outlier_eps[*i*].txt in the ./statistics/text where *i* indicates the month.

   For example, outlier_eps1.txt is the text file that contains epsilon outliers for January.

4. Create epsilon surfaces using the Geostatistical Analyst in ArcMap.

   1) Click "rainfall disaggregation" button in the menu bar. This tool was created using VBA code to select the rain gauge features used for creating epsilon surface by the Geostatistical Analyst. The following steps show how to add the tool in ArcMap tool bars.

      (1) Right-click on any tool bar and select customize…

      (2) From the Commands tab, scroll to the UIControls category.

      (3) Click on the New UIControl button. Choose UIButtonControl as the UIControl Type, and click Create.

(4) project.UIButtonControl1 will be shown in the new control list. Rename it as "project.populate". Drag this new tool onto any tool bar. Change the name of tool as the "rainfall disaggregation".

(5) Right click on the tool and click View source. Two lines of code (the Private Sub and End Sub) are added. This event handler will have the name "populate_click". Put the function "frmPopulateStats.Show" into the event handler. When the tool is clicked, the event handler will be executed.

2) Click "Apply" to select the rain gauge features to disaggregate

This reads rzopt_eps.out and outliers_eps[i].txt file, and selects the features that are not the outliers in the rain gauge stations feature layer in ArcMap.

5. Run Geostatistics Analyst in ArcMap

1) Determine the coefficient of Box-Cox transformations based on the statistics (kurtosis, mean and median) by displaying histogram, and Q-Q plot in the Geostatistical Analyst

2) Run Geostatistical Analyst and create epsilon surfaces

3) Convert Geostatistical Analyst layer to grid layer (use default size)

4) Create rzGAEpsilon.dat that contains interpolated epsilons at the rain gauge stations by @mRainDisaggregation.writeEpsilonSurfaceValues

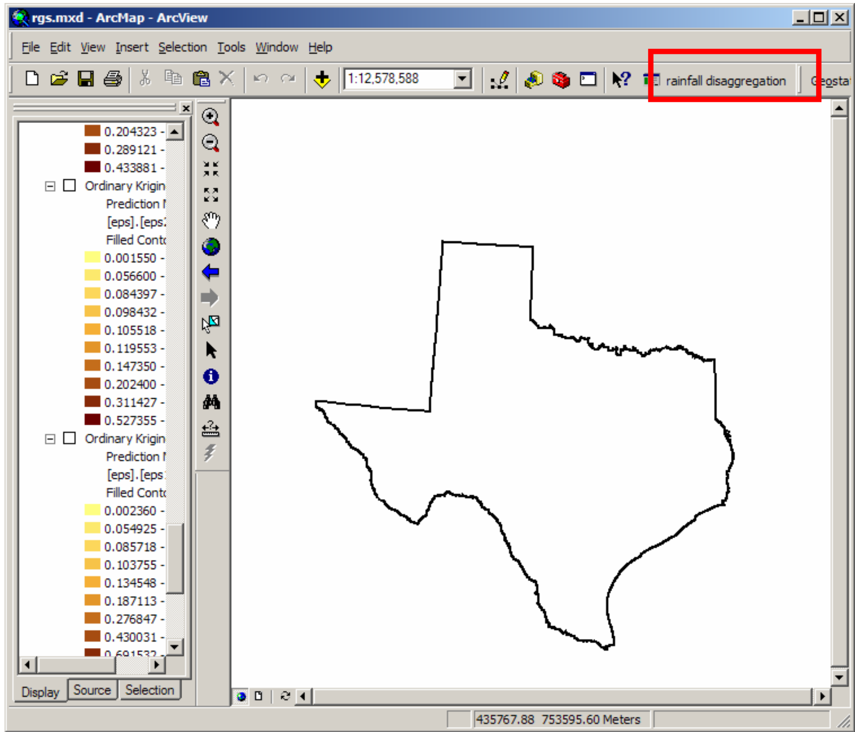


Figure A-15. Rainfall disaggregation button in the red box.

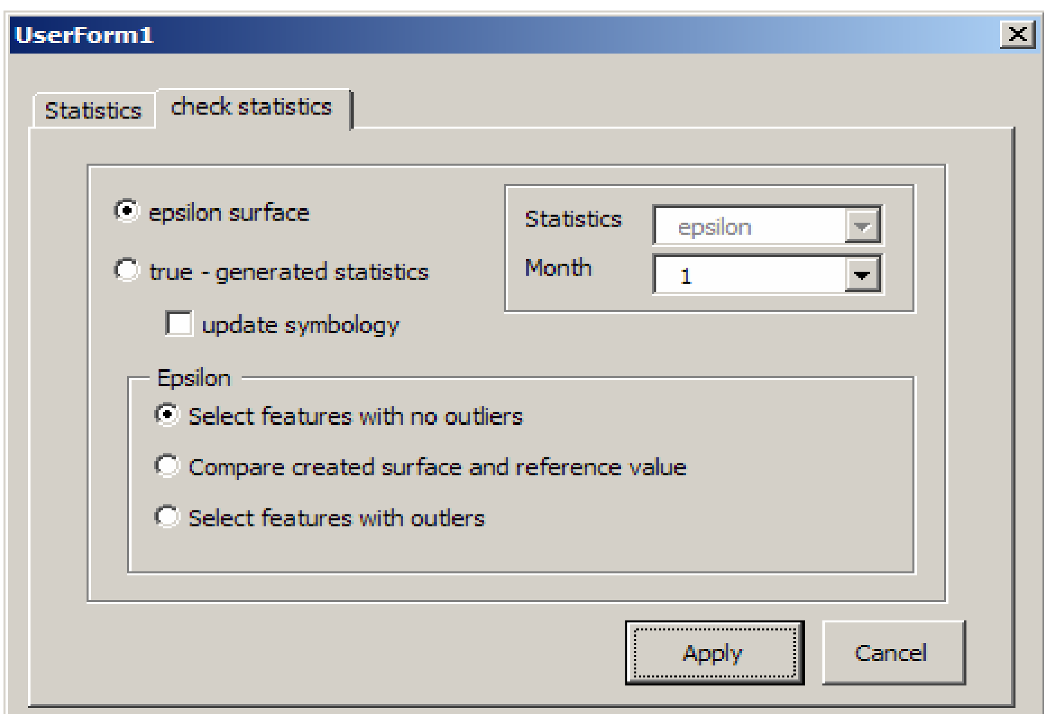


Figure A-16. Userform that allows the user to select month for epsilon surface.

A.2.6. Add X, Y Coordinates in ArcMap

In order to create ε surface, ArcGIS Geostatistical Analyst is used. ArcGIS Geostatistical Analyst is an extension of ArcGIS desktop (ArcView, ArcEditor, and ArcInfo) that provides tools for analyzing spatial data. The Geostatistical Analyst requires the information of the text file "TX.dat" that contains raw rainfall data (NOAA gauge number, start time in month day year, stop time in month day year, latitude and longitude in degrees minutes seconds, and latitude and longitude in decimal degree). But use of the data as input to Geostatistical Analyst requires the data in a specific ESRI binary data format (e.g., point featureclass). The data in delimited text files can be transformed to point feature class using the functionality of ArcMap. The files with extensions (.txt, .asc. csv. .tab. dbf) can be imported into ArcMap. ArcMap allow the user to directly access these types of data and work with them as tables. Thus, TX.dat needs to be converted to one of these files. By default, files with a .txt, .asc or csv extension are interpreted as comma delimited while files with a .tab extension are interpreted as tab delimited. TX.dat contains the information of latitude and longitude of 646 rain gauges in degrees, minutes, and seconds, which can be converted to decimal degrees and stored in a numeric field.

1. Import XY (Point) Data from MS Excel to ArcMap

The following outlines the steps.

1. Import a TX.dat into MS Excel and convert degrees/minutes/seconds to decimal degrees and save it as dBase
2. Import dBase table of XY data into ArcMap, and specify the geographic coordinate system

2. Import a TX.dat into MS Excel and Convert Degrees/minutes/seconds to Decimal Degrees

1. Open TX.dat with MS Excel.
2. Choose Data -> Text to Columns
3. Check original data type as delimited and click next
4. Check space as delimiters and click finish.
5. Type in the following new column headings

   (Lat_deg, Lat_min, Lat_sec, Long_deg, Long_min, Long_sec, Lat_dec, Long_dec)

6. Populate Lat_dec and Long_dec as follows

   Long_dec =-(Long_deg+(Long_min+Long_sec/60)/60)

   Lat_dec = -(Lat_deg+(Lat_min+Lat_sec/60)/60)

7. Save TX.dat as TX.dbf (dBase IV).

3. Import dBase Table of XY Data into ArcMap, and Specify the Coordinate System

1. Open ArcMap with a new empty map document
2. Choose tools -> ADD XY DATA…
3. Locate the directory in which the table is stored
4. Select TX.dbf and click ADD
5. Specify the X field as Long_dec and the Y field as Lat_dec
6. Click the Edit button to specify the spatial reference

7. Click on the SELECT button for a predefined coordinate system

8. Choose Geographic Coordinate System -> North America -> North America Datum 1983

9. Click Ok

10. Double-click on the Layers in the Table of Contents (TOC) and select button for a predefined coordinate system

11. Choose Project Coordinate System -> Continental -> North America -> USA_Contiguous_Albers_Equal_Area_Conic_USGS_version

12. Save as txrgs in the rain disaggregation.mdb

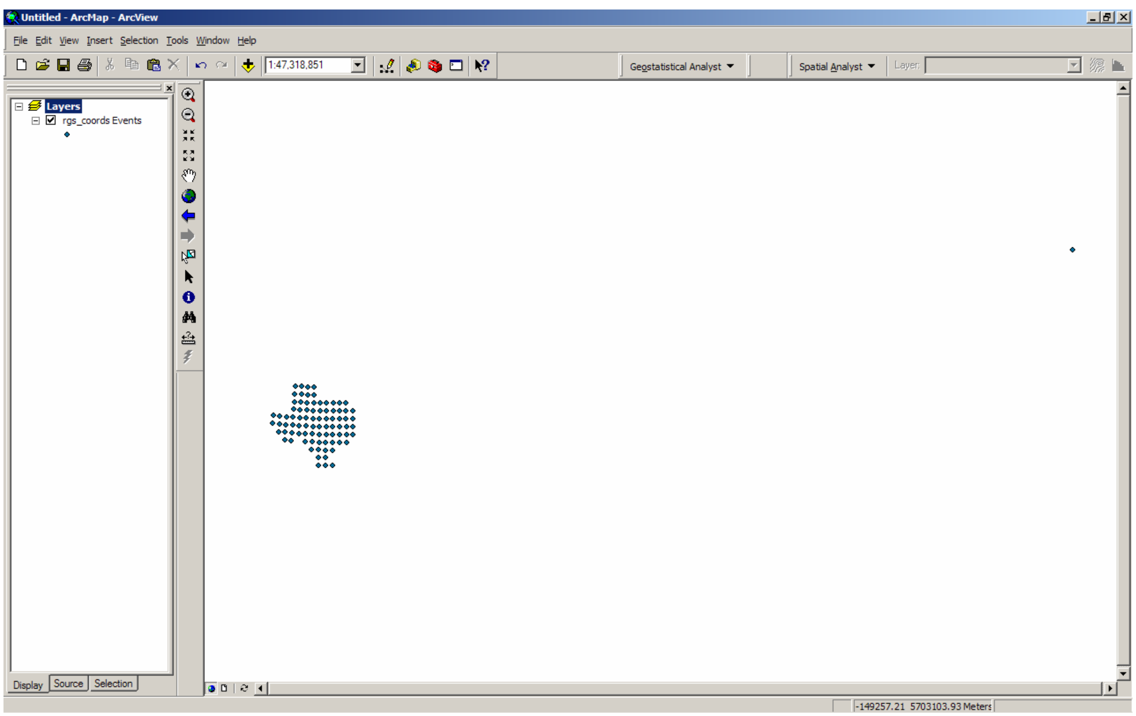


Figure A-17. Display of the feature class of rain gauges.

Out of 646 rain gauges, 115 (18%) were located far outside Texas (Figure A-17). These data were considered to be invalid and removed. The remaining 531 rain gauges were used for this study. The txrgs feature class consists of 19 fields (objectid (1), shape*(1), user-defined gauge number(1), NOAA gauge number(1), start time in month day year(3), stop time in month day year(3), latitude and longitude in degrees minutes seconds(6), and latitude and longitude in decimal degree(2)) where (#) indicates the number of fields used, for example, start time in month day year takes 3 fields. The period of record for the rain gauge stations is needed to filter out the rain gauges of few years of record. Thus, a field called "recYear" was added to txrgs feature class using @mdisaggregation.AddNewFields function. The field was populated using the function @mdisaggregation.popYearDiffField. The txrgs feature class is the base feature class containing all the necessary information about the 531 rain gauge stations except rain depths. In order to generate epsilon surfaces, 12 epsilon fields corresponding to 12 months need to be added to the txrgs feature class, and the feature class becomes larger in size. Rather than adding more fields to the txrgs feature class, another feature class was created to contain this information (e.g., user-defined gauge number, NOAA gauge number, recYear, epsilons). In order to do that, the txrgs feature class was exported to raindisaggr_eps.mdb and saved as a point feature class named "eps".

Once the eps feature class was created, the Geostatistical Analyst in ArcGIS was used to create epsilon surfaces using spatial interpolation techniques. There are two groups of interpolation techniques: deterministic and geostatistical. The Geostatistical Analyst has several deterministic techniques (e.g., inverse distance weighting, and global polynomial) and geostatistical interpolation methods (e.g., Kriging, and Cokriging).

In this dissertation, ordinary kriging method was applied. Before creating ε surfaces using the ordinary kriging technique, it is necessary to investigate whether the data is normally distributed since the ordinary kriging method requires that the data is normally distributed. When the data is skewed, the data is required to be transformed to make it normal. Before the normalization is performed, trend is required to be removed to justify assumptions of normality and stationary after the transformation is conducted. There are two options (e.g., Box-Cox transformations and arcsine transformations) for the transformation in the ordinary kriging.

Outliers can be identified by using the histogram tool in the Geostatistical Analyst. Unusually high data were found to be scattered around Texas, and surrounded by very different values. In such case, the unusual data were considered outliers and removed. The data were found to be positively skewed and a Box-Cox transformation, a built-in function in Geostatistical analysis, was not successful since kurtosis was greater than 3 (leptokurtic). An alternative way to obtain normal distribution is to use area or score transformation which is a normal score transformation methods. The area transformation was conducted using the function $main_IdentifyEpsilonOultiers in the SPSS script file called dsgtool.SBS which is stored in the directory ./scripts. The $main_IdentifyEpsilonOutliers also created text files named outlier_eps[#].txt where [#] indicates the numeric value of month (i.e., January=1, December=12). This text files were read by the disaggregation tool shown in Figure A-16 to select the features only used for spatial interpolation.

Once the outliers in the epsilon values were identified and removed, the Box-Cox transformation was applied to normalize the epsilon values to utilize the ordinary kriging in the Geostatistical Analyst. The ordinary kriging creates epsilon surfaces in the form of GA (Geostatistical Analyst) layers in the ArcMap. GA layers, however, do not allow the user to automatically extract values of at given location. Thus, the GA layers were converted to grids to which it was possible to access programmatically. The interpolated epsilon values were extracted from the grid layers and written in a text format by @mRainDisaggregation.writeEpsilonSurface Values. The interpolated epsilon values could be used to verify the presented disaggregation model.

A.2.7. Data Manipulation

In this disaggregation method, recorded hourly rainfall data were used. Daily rainfall was obtained by accumulating hourly rainfall data for 24 hours. Synthetic hourly rainfall was determined by disaggregation of daily rainfall. Their corresponding statistics are called true (or observed) statistics and simulated statistics. The text file stats.dat stored in the directory ./results/tx/ contains simulated statistics obtained by 531 gauges. The file contains 281,961 (531×531) rows for each simulation. The columns are daily gauge, disaggregating gauge, and 12 columns each for epsilon, probability of zero rainfall (p_o), average (m), variance (σ^2), and 1-hr lag autocorrelation (ρ). The average is the total rainfall divided by total number of hours for a specific month whether raining or not. The variance is hourly rainfall variance for a specific month. The stats.dat file is too large in size to efficiently handle using MATLAB. It is necessary to split stats.dat into several MATLAB format files (.mat). The following MATLAB code reads stats.dat and saves it as allstats.mat using a MATLAB built-in function textread, and a user-defined MATLAB function mxsave. Mat-files (.mat) are the binary MATLAB format files which are used for saving data. The mat files are easily converted to a matrix by using the built-in MATLAB function load or user-defined function mxload. Those functions can be executed in MATLAB command window as follows.

>> allstats = textread('stats.dat', '','whitespace','\b');

>> mxsave(allstats, ‘allstats’, ‘allstats’)

The function mxwstats.m was created to divide the allstats.mat into 48 MATLAB files (12 months × 4 statistics) and stored in the directory ./storm/dsg/data. Each MATLAB file contains 531 rows and 531 columns (i.e., 531×531 matrix). The input argument for the function mxwstats.m is the allstats.mat. The MATLAB commands to run the associated functions are as follows.

>>allstats = mxload('.', allstats)

>> *mxwstats(allstats)*

The columns are the disaggregating gauges, and the rows are the disaggregated gauges for a specific statistic and month. The diagonal elements in the matrix indicate auto-disaggregation statistics (statistics of the disaggregated rainfall data by its own database).

The true_stat.dat contains the true statistics (measured statistics). The true statistics contains one row per station. Columns 1 and 2 are both the station number. The remaining statistics are for the measured precipitation data:

3-14 is the p0,

15-26 is the average rainfall,

27-38 is the variance

39-50 is the lag-one autocorrelation coefficient.

This file also was converted to a MATLAB file (truestats.mat) and stored in the directory ./storm/dsg/data.

A.2.8. Performance Statistics for Auto-Disaggregation

To evaluate the auto-disaggregation statistics, the statistics of auto-disaggregation time series at each gauge station were compared with the true statistics. The error statistics (measured – simulated statistics) were calculated and saved by the function &mxprnerrstats.m. The &mxprnerrstats.m creates 48 MATLAB data files corresponding to 4 statistics and 12 months. Each file contains 531 rows and 531 columns (a 531×531 matrix). Each row corresponds to a gauge to be disaggregated. Each column corresponds to the disaggregating gauge. The file structure of error statistics is same as that of simulated statistics. Originally the error statistics were used to find spatial patterns of database used for disaggregation. The MATLAB function &main_modperf.m is the main function which includes a set of functions associated with the evaluation of auto-disaggregation performance.

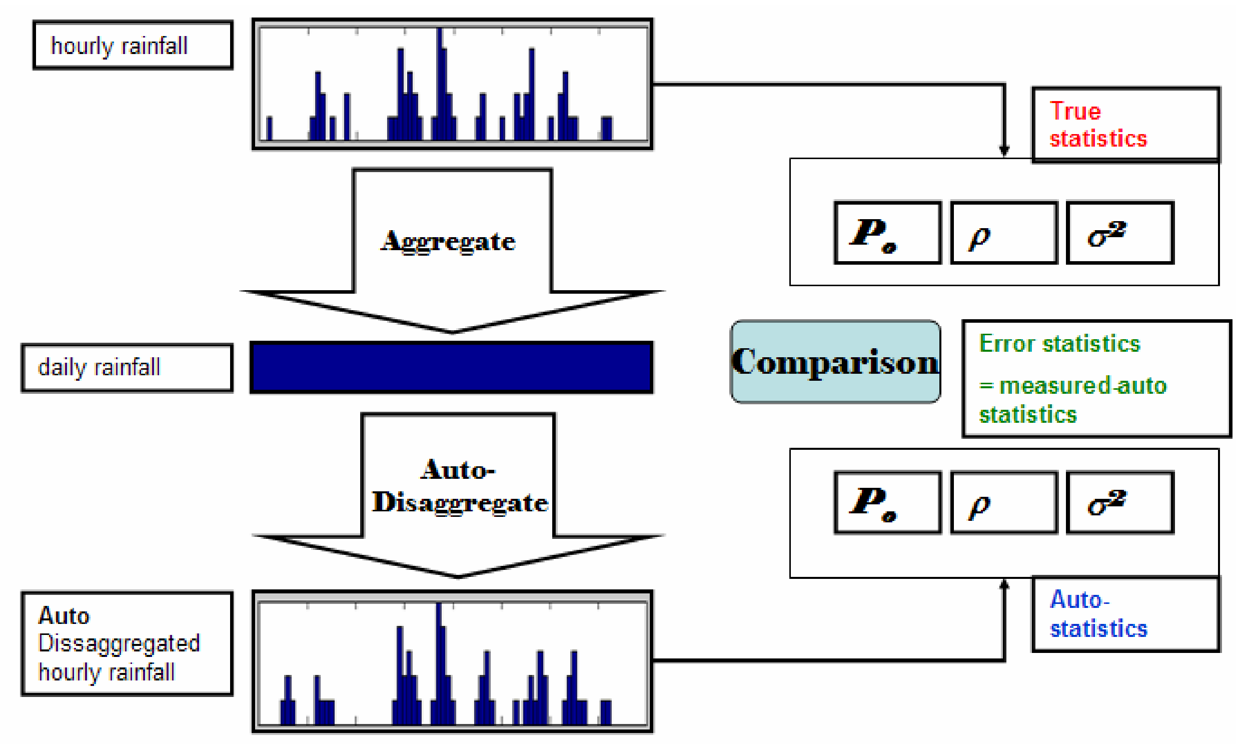


Figure A-18. Comparison of measured and auto-disaggregation statistics.

A.2.9. Gauge Performance Trends

The regional trend analysis is accomplished by a MATLAB function &main_regionaltrend.m. The &main_regionaltrend.m contains a set of functions (e.g., &mxgems.m, &mxmixdenplot.m, and &mxtitlepub.m) associated with it.

The gauge characteristic trend analysis is accomplished by a MATLAB function &main_gct.m which contains a set of functions (e.g., &mxgems, &mxgtms, &mxfid, and &mxmixdenplot) associated with it.

The proximity trend analysis is accomplished by a MATLAB function &main_proximity.m.

Verification of Disaggregation Method Across Texas

A set of 23 randomly-selected rain gauges were used for verification of disaggregation method using the state of Texas storm database. Verification of disaggregation method for specific gauges can be accomplished by a MATLAB function &main_modver.m. Comparison between general disaggregation and true statistics using the state of Texas storm database can be accomplished by a MATLAB function &main_modperf.m. The &main_modpef.m returns performance measures between the observed and disaggregated data for the verification. The performance measures are the statistical summaries, MAEs, RMSEs, correlation coefficients, and the degrees of agreement of the observed and disaggregated data.

A.3. STORM FEATURE IDENTIFICATION AND TRACKING ALGORITHM

The programs used for the storm feature identification and tracking algorithm are DOS, Cygwin, and MATLAB 7.0.

A.3.1. Directory Structure

All the input, output and script files for the analysis are stored in top-level directories and its subdirectories. The top-level directory is a working directory that contains the necessary files and subdirectories. The top-level directory can be managed under a project directory. There are three main project directories. One is the directory for running the programs and storing the inputs, outputs, and associated files of MATLAB and the second one is the directory which contains the inputs, outputs, and associated files of C, Cygwin, and DOS program. The last one is the directory which contains the executable files to convert NEXRAD MPE data format. The directories are listed below.

1. Top-level directories for C, Cygwin, and DOS program

./	Contains subdirectories and text files (.txt and .out)
./program	Contains subdirectories, there are c source codes, and readme text files (.txt) for explaining the usage of the c functions
./xmrg_2003_BE	Binary data files for NEXRAD MPE of 2003 which are formatted on Big Endian machine
./xmrg_2003_LE	Binary data files for NEXRAD MPE of 2003 which are formatted on Little Endian machine

Subdirectories

The section provides information about some of the subdirectories and their contents.

./program

./source	Files that contains C source codes (.c)
./data	Test data for the c programs
./original	Directory that contains the original source code downloaded from a website

2. Top-level directory for the executable files to convert NEXRAD MPE data format (c:/Cygwin/bin)

./cygwin/bin	Executable files

3. Top-level directories for MATLAB (c:/MATLAB7/work/toolbox/)

../	Common MATLAB m-files (.m) for the three methods (the derivation of ARFs, the disaggregation method, and the storm feature identification and tracking method)
./	MATLAB m-files (.m) shared by all the tools under /toolbox
./statlearn	MATLAB objects and functions modeling multivariate densities and classification procedures
./stix	functions for statistical analysis
./storm	subdirectories are related to the methods and common functions (.m) shared by the methods

./statlearn

./unsupervised	MATLAB objects (density, gaussdens, diagdens, balldens, pca-dens, mixdens, unifdens, and mxmoddens)

./storm

./dsg	subdirectories and functions associated with the rainfall disaggregation
./sfit	subdirectories and functions associated with the storm feature identification and tracking algorithm

./storm/dsg

./	MATLAB script files (.m) and MATLAB figure files (.eps and .fig)
./data	MATLAB data files (.mat) for observed statistics and error statistics

./storm/sfit

./	MATLAB script files (.m) and MATLAB figure files (.eps and .fig)
./data	MATLAB data files (.mat) for storm feature objects and storm feature paths

A.3.2. Data Preparation

Four steps were taken to convert the NEXRAD MPE data to MATLAB data files which are the standard data form in MATLAB.

1. Download the archived NEXRAD MPE data and the associated C programs from the website

2. Extract the archived NEXRAD MPE data with unzipping program

3. Convert the extracted binary files formatted on Big Endian machines into binary files on Little Endian machines (e.g., Windows XP on Intel machine).
4. Convert the binary files on Little Endian machines into the MATLAB data files (.mat) which is a matrix in MATLAB.

The NEXRAD MPE was collected from the River Forecast Centers (RFCs) on a daily basis. The daily files are constructed and archived as monthly files via the UNIX 'tar' utility. The NEXRAD MPE grids are stored in a binary file format called XMRG. The NEXRAD MPE for WGRFC can be downloaded at http://dipper.nws.noaa.gov/hdsb/data/nexrad/wgrfc_mpe.php. Three C programs (read_xmrg.c, read_xmrg_lin.c and reverse_byte_ordere.c) that read an XMRG file are necessary to convert the downloaded binary files into ASCII files. Those can be downloaded at http://www.nws.noaa.gov/oh/hrl/dmip/nexrad.html. In order to handle the archived binary files and to run C programs, Cygwin is installed and used. Cygwin is a collection of free software tools originally developed by Cygnus Solutions to provide Linux-like environment for Windows. The program can be downloaded for free at http://www.cygwin.com. The read_xmrg.c program reads XMRG file on Big Endian machines (e.g., HP UNIX workstations). On the other hand, the read_xmrg_lin.c and reverse_byte_order.c program read XMRG files on Little Endian machines (e.g., Intel machine). The following description about the C program reverse_byte_order.c is excerpted from the comments in the source code of the C program.

> This reverse_byte_order.c reverses the ordering of the bytes in each 4-byte word of an integer array. For example consider the following
>
> 4-byte word whose bytes contain the characters 'A', 'B', 'C', and 'D':
>
> byte 1   byte 2   byte 3   byte 4

A B C D

This routine will reverse the ordering of these bytes in this 4-byte word so that the contents of each byte of the word will appear as follows:

byte 1 byte 2 byte 3 byte 4

D C B A

The need for this routine arises from differences in memory architecture across different computer platforms. The two memory configurations that need to be accommodated are the "Big Endian" and "Little Endian" architectures. In the "Big Endian" architecture, the left-most byte in a word is the most significant byte. In the "Little Endian" architecture, the right-most byte in a word is the most significant byte. As another example, consider a 4-byte integer which contains the value 66. On a "Big Endian" system, the binary pattern in a word would appear as follows:

00000000 00000000 00000000 01000010

On a "Little Endian" system, the binary pattern would appear as follows:

01000010 00000000 00000000 00000000

This routine ensures that a GRIB2 message will be accurately decoded regardless of the memory architecture of the computer that it is being decoded on.

New C programs (gen_xmrg_bin_win.c, mx_gen_xmrg_bin_win.c, and mx_gen_xmrg_bin_win.exe) were developed to modify the binary files which were formatted on Big Endian program to the binary files that were formatted on Little Endian machines. The binary files on Little Endian machines do not have header information of the original binary files (downloaded from the website) so that this can reduce the file size, and speed up the data processing. The mx_gen_xmrg_bin_win.c and gen_xmrg_bin_win.c were compiled to create the executable file mx_gen_xmrg_bin_win.exe and stored in the ./Cygwin/bin. The following script on Cygwin command prompt is used to create mx_gen_xmrg_bin_win.exe.

At the Cygwin prompt,

cc -o mx_gen_xmrg_bin_win -mno-cygwin mx_gen_xmrg_bin_win.c gen_xmrg_bin_win.c

Once the executable file was created, the binary files on Little Endian machines were created on Cygwin command prompt by the mx_gen_xmrg_bin_win.exe as follows.

At the Cygwin prompt,

mx_gen_xmrg_bin_win.exe arg1

where arg1 is an input argument for the mx_gen_xmrg_bin_win.exe. The input argument (arg1) is the name of a text file which contains the list of names of the binary files on Big Endian machine. The text file were created by the following dos command.

At the DOS prompt,

C:\NEXRAD_DATA\xmrg_2003_BE> dir /B >>../ xmrgBE_2003_fnl.txt

The binary files formatted on Little Endian machine were also created in the same directory. These files were moved into the data directory (./xmrg_2003_LE) for the better file management. The storm feature identification and tracking algorithm was developed in MATLAB. For that reason, the binary files were reformatted into MATLAB data file so that it was efficiently handled by MATLAB functions and scripts. A set of MATLAB functions (&mxbin2mat.m, &mxreadxmrg.m, and others) was used to read the binary files and saved those as the MATLAB data files. The &mxbin2mat.m is the main function that calls the associated functions. The input argument to the &mxbin2mat is the name of the text file that contains the list of file names of the binary files. The text file that contains the list of file names of the binary files was created by the following DOS command.

At the DOS prompt,

C:\NEXRAD_DATA\xmrg_2003_LE> dir /B >>../ xmrg_2003_le_fnl.txt

The MATLAB files were created from the binary files on Little Endian machine by the following MATLAB script.

At the MATLAB command window,

>> mxbin2mat xmrg_2003_le_fnl.txt

The created MATLAB data file is the matrix of size 329 × 420 which is the exact same dimension of the WGRFC NEXRAD grid. The created MATLAB data files were moved to the directory ./storm/sfit/data/nxmpe2003/. A text file was created to contain the list of the MATLAB data file names that was used to manage the files. The list is useful to fast search and find the file when it is needed as an input to image analysis. The following dos command was used to create the list of the NEXRAD MPE file names.

At the DOS prompt,

C:\MATLAB7\work\toolbox\storm\sfit\data\nxmpe2003> dir /B >>../ nxmpe2003fnl.txt

The created text file was converted and saved to a MATLAB data file nxmpe2003fnl.mat which is a cell array of size n ×1 where n is the number of the MATLAB data files. The conversion can be done with the following MATLAB script.

At the MATLAB command window,

>>nxmpe2003fnl= textread('nxmpe2003fnl.txt', '%s');

>>save nxmpe2003fnl ../nxmpe2003fnl;

The following script was used to load nxmpe2003fnl (a workspace variable) from disk. The nxmpe2003fnl is a matrix of size $n \times 1$ where n is the number of rows.

At the MATLAB command window,

>>nxmpe2003fnl = mxload('../', 'nxmpe2003fnl');

Figure A-19 shows the nxmpe2003fnl workspace variable which is a matrix of size 8613×1.

	1
1	Xmrg_01012003_01z_WG_win
2	Xmrg_01012003_02z_WG_win
3	Xmrg_01012003_03z_WG_win
4	Xmrg_01012003_04z_WG_win
5	Xmrg_01012003_05z_WG_win
...	...
...	...
124	Xmrg_02032003_07z_WG_win
125	...
126	Xmrg_12302003_23z_WG_win

Figure A-19. nxmpe2003fnl workspace variable: gray area is the indexing of rows and columns for the matrix which is not included in nxmpe2003fnl.

A.3.3. Storm Identification and Tracking Algorithms

The storm feature identification and tracking algorithm consists of three main MATLAB toolboxes: sfit, statlearn, and stix. The sfit toolbox was developed as part of this work. The statlearn toolbox was originally developed by Guillaume Bouchard and modified by part of this work. The stix toolbox is developed by Anders Holtsberg. Both statlearn and stix are public domain software. The MATLAB functions shared by the MATLAB toolboxes are located at the C:\MATLAB7\work, C:\MATLAB7\work\toolbox, and C:\MATLAB7\work\toolbox\storm. The NEXRAD MPE data used for the identification and tracking algorithm was stored at the C:\MATLAB7\work\toolbox\storm\sfit\data\nxmpe2003.

A.3.4. Data preparation for Brazos County

Storm events occurred over Brazos County were extracted by the function &mxcheckrainyday.m. Storm event is defined as consecutive NEXRAD precipitation images that have rainfall over the Brazos County. The NEXRAD precipitation images were archived in the form of MATLAB data file (.mat). A NEXRAD image or consecutive NEXRAD images that have no rainfall over the Brazos County break the series of NEXRAD images into storm events. Storm feature is an identity (an irregular shaped polygon) that is a fully segmented object obtained from a NEXRAD image in the process of storm feature identification. The function &mxcheckrainyday clipped the polygon area of Brazos County from the NEXRAD precipitation image. It checks the cells in which rainfall depth is greater than 0. The &mxcheckrainyday returned a matrix of size $n \times 1$ where n is the number of NEXRAD images. The matrix contains the number of cells that rainfall occurred over the Brazos County from a NEXRAD image. Storm events were identified by the function &mxcreaterset. The identified storm events were stored in a matrix (a MATLAB variable). The matrix is specifically referred to as a record set for the storm events. The matrix consists of rows and columns. The rows can be referred to as records. The columns can be referred to as fields. The record set for the storm events was created by the function &mxcreaterset by passing the output of the &mxcheckrainyday into input argument to the &mxcreaterset. Each record (or row) in the record set indicates a single storm event. The &mxcreaterset returns a matrix that consists of three columns and n rows where n is the number of the records. Each row corresponds to storm events. Three columns are the record set id, the row index of the nxmpe2003fnl.mat, and the number of consecutive NEXRAD images. The first column is the record set id (RID). The second column of the record set corresponds to the row index of the nxmpe2003fnl.mat (RIDX). The starting date of a storm event can be obtained from the nxmpe2003fnl.mat by finding the same row index as the value of the second column of the record set. The third column indicates the consecutive NEXRAD images that have rainfall over the Brazos County.

	RID	RIDX	#Images
	1	2	3
1	1	260	15
2	2	276	13
3	3	609	8
4	4	800	2
5	5	877	6
...	...	...	...
...	...	...	...
124	124	8403	2
125	125	8532	3
126	126	8540	5

Figure A-20. Record set of storm events (gray area is the indexing of rows and columns of the matrix, blue area is the field names, and white area is the MATLAB variable).

A.3.5. Setup for Environment Variables for MATLAB

The environmental variables for the identification and tracking program are initialized by the MATLAB script &mxinivar_sfit located in the directory c:/MATLAB/work/. The &minivar_sfit sets up the working directories by adding paths to the environment variables of MATLAB. The paths are the directories of the source codes and the data files. The &minivar_sfit also loads the saved MATLAB data (.mat) into the memory (assigning the data to the associated MATLAB variables) so that the saved MATLAB data are ready to use in application.

A.3.6. Main Script Files for Storm Feature Identification and Tracking

The main script file for storm feature identification and tracking is &the task_ssit.m which is the workspace for setting up all the necessary model parameters and running all the functions associated with storm feature identification and tracking. It identifies all the storm features in a NEXRAD image. A storm feature is stored as a structure array in MATLAB. Structure array in MATLAB is a collection of records with the specified fields and values. Likewise, the storm features can be considered as structure arrays. The structure array is embedded in a cell array as an element. Cell array in MATLAB is a container for any type of MATLAB data, including other cells. Thus, the storm features in a NEXRAD image are stored as a single cell array which contains structures as elements. Figure A-21 shows the schematic diagram of how storm features are stored in MATLAB variables. The record in pink color in the record set of storm events has two consecutive NEXRAD images, occurred from 7:00 AM to 8:00 AM on Feb. 3rd in year 2003. In the first NEXRAD image (tagged as 02032003_07), four storm features are identified, and each is stored as a single structure array. The four structure arrays are embedded in the single cell array as shown in the figure. The second NEXRAD image (tagged as 02032003_08) has three storm features identified, and each is stored as a single structure variable. The three structure variables are, then, embedded in a single cell array. Two cell arrays are again embedded in another single cell array for the record. The cell array is hereafter referred to as storm event object or record object. The structure array for storm feature is hereafter referred to as storm feature object. The record object is saved as a MATLAB data file named oc4.mat in the directory ./sfit/data. Likewise, all records are saved as oc[RID].mat where the [RID] is the numeric value of the RID field, and [RID] is replaced by the corresponding value when it is applied.

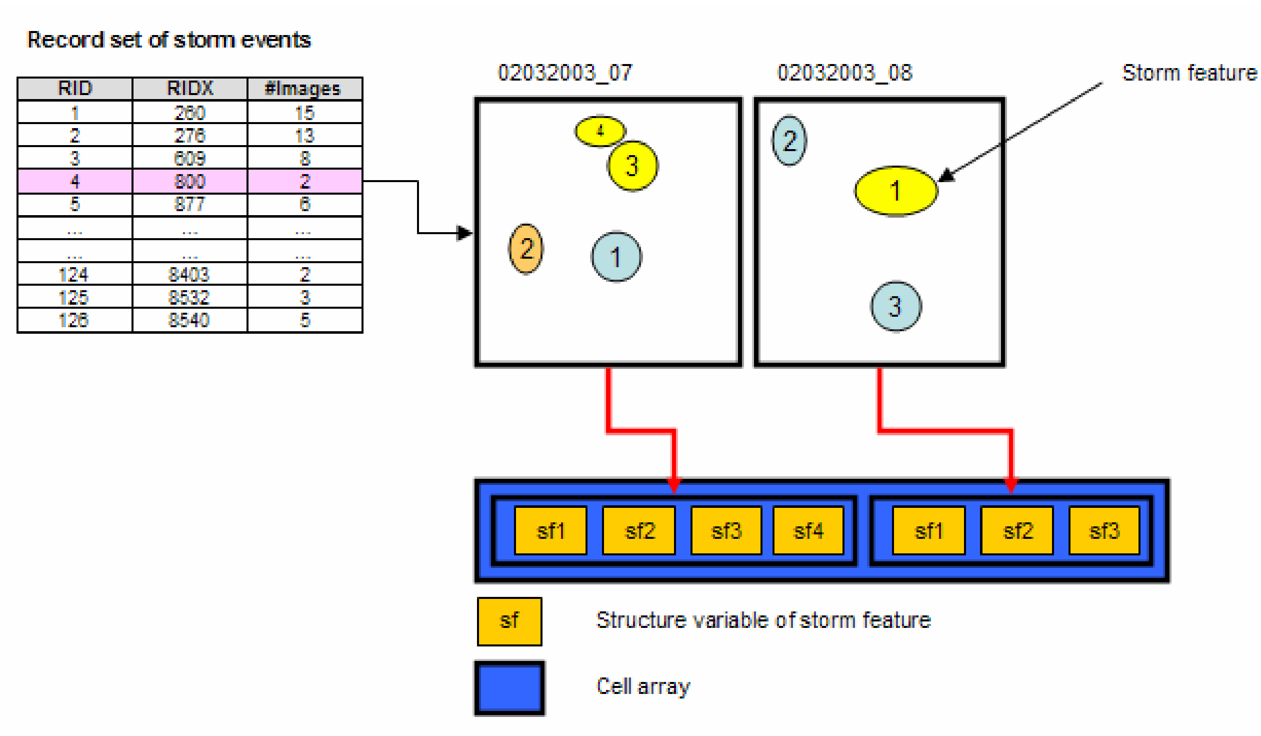


Figure A-21. Schematic diagram of structure and cell array of storing storm features.

Figure A-22 shows the different types of MATLAB variables in different colors. Figure A-23 shows the structure variable of storm feature with fields and brief explanation. Figure A-24, Figure A-25, and Figure A-26 show the object variables defined in the statlearn toolbox.

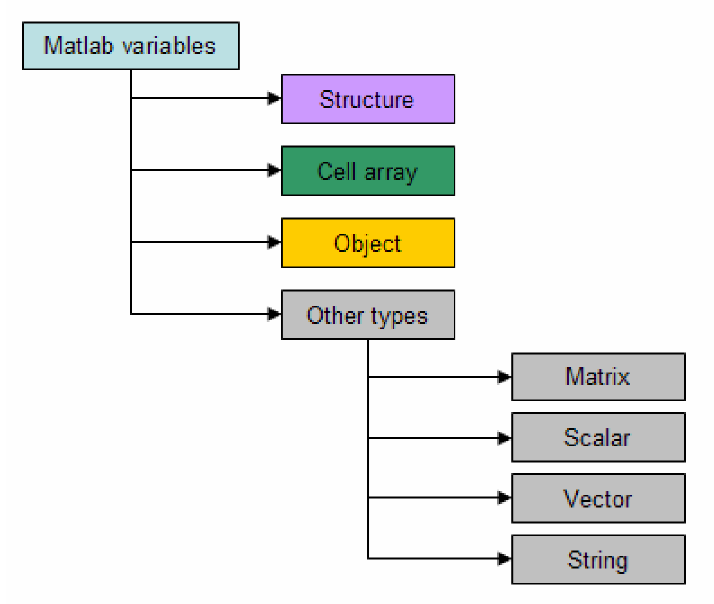


Figure A-22. MATLAB variables in different colors.

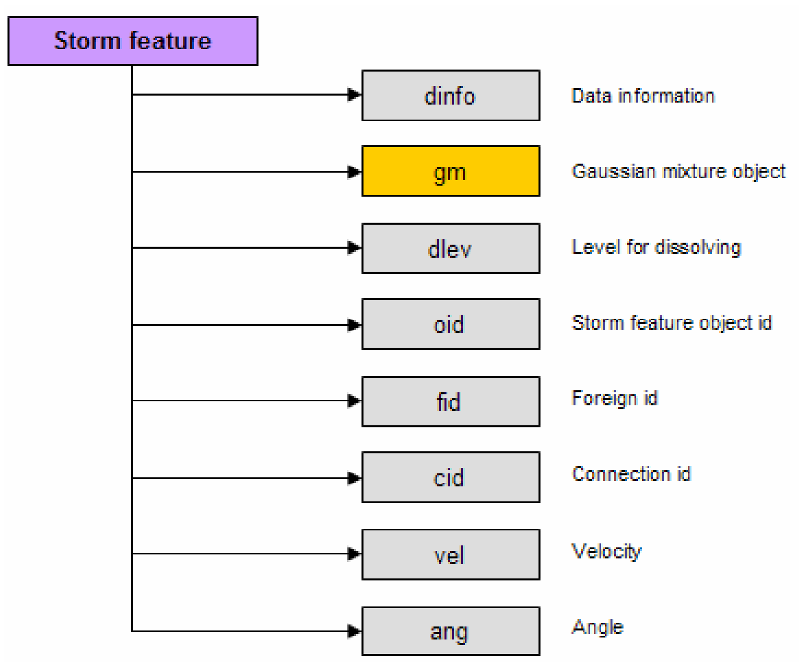


Figure A-23. Structure array for storm feature (storm feature object).

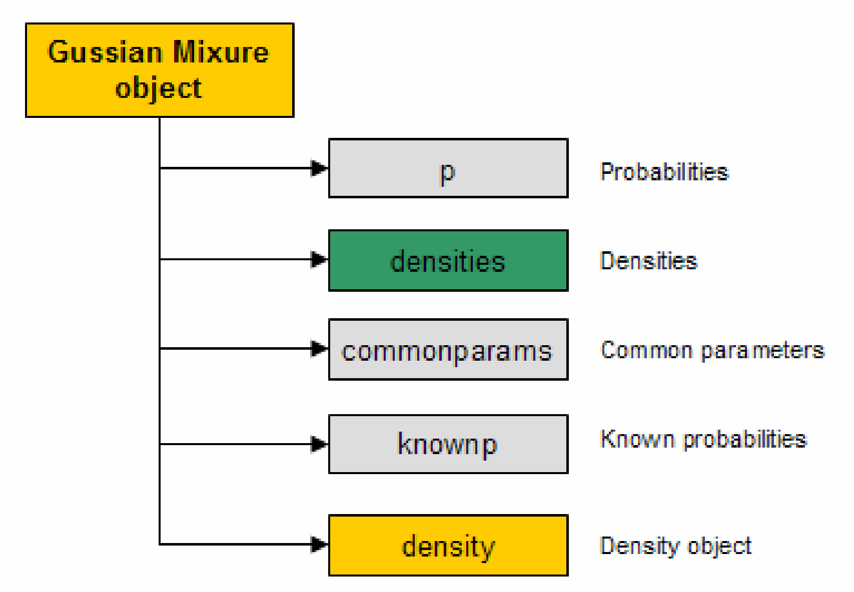


Figure A-24. Gaussian mixture object variable.

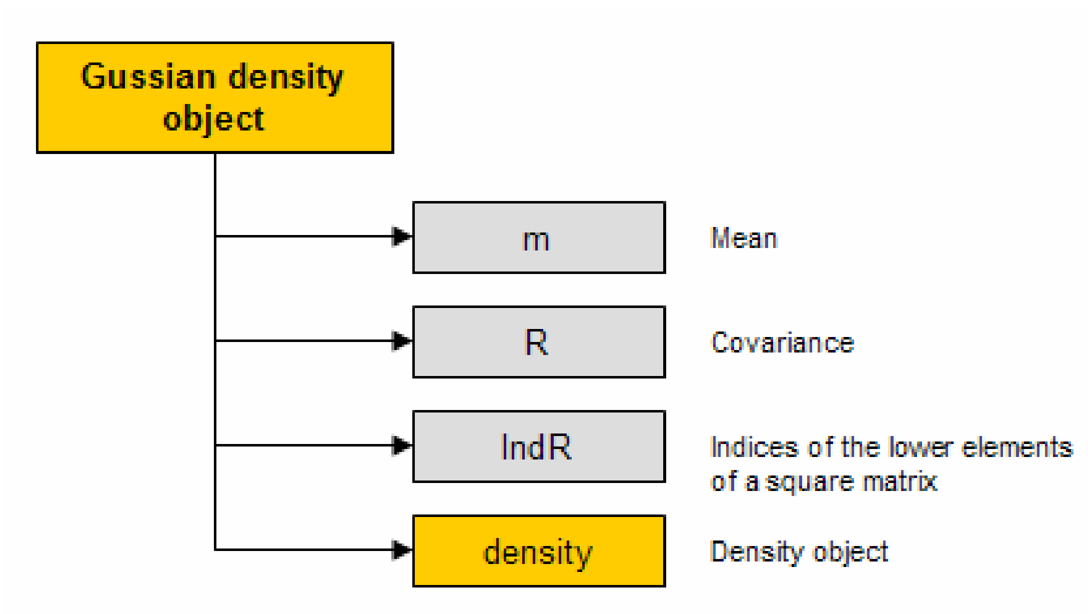


Figure A-25. Gaussian density object variable.

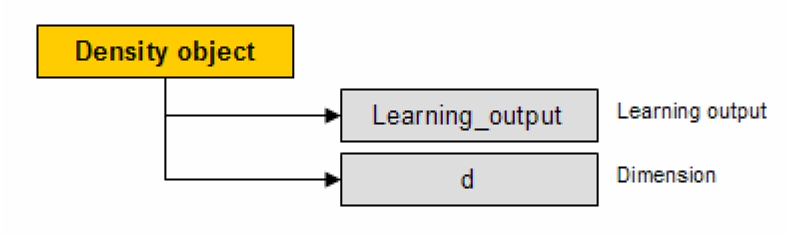


Figure A-26. Density object variable.

Looping through the record set of storm events, the &task_ssit reads each record at each iteration. Based on the information of the record, the &task_ssit loads the corresponding NEXRAD images into MATLAB variables. The &task_ssit calls the function &mxgmssit and passes all the model parameters and the NEXRAD images associated with the record into the function as input arguments. The &mxgmssit consists of two main algorithms: storm feature identification and storm feature matching. The &mxgmssit.m calls the function &mxidse for storm feature identification. The &mxidse identifies all the storm features from the given NEXRAD images. All the identified features are stored as storm feature objects, and then embedded in a cell array. Once the storm features are identified and stored as MATLAB variables, a given storm feature is matched with the ensuing storm feature by the function &mxgmmatch. The input argument for the &mxgmmatch is a set of model parameters associated with the tracking method, the record object, and two consecutive images, called an image pair. For example, *n* NEXRAD images have *n*-1 image pairs. The cid, fid, velocity and angle field of the storm feature object are calculated based on the relationships of two matching storm features. If there is no matching storm feature, the velocity and angle field of the given storm feature have empty value as a default. Figure A-27 shows that how storm feature objects are matched and their properties can be calculated. The &task_ssit saves the record object in the data directory for each iteration process.

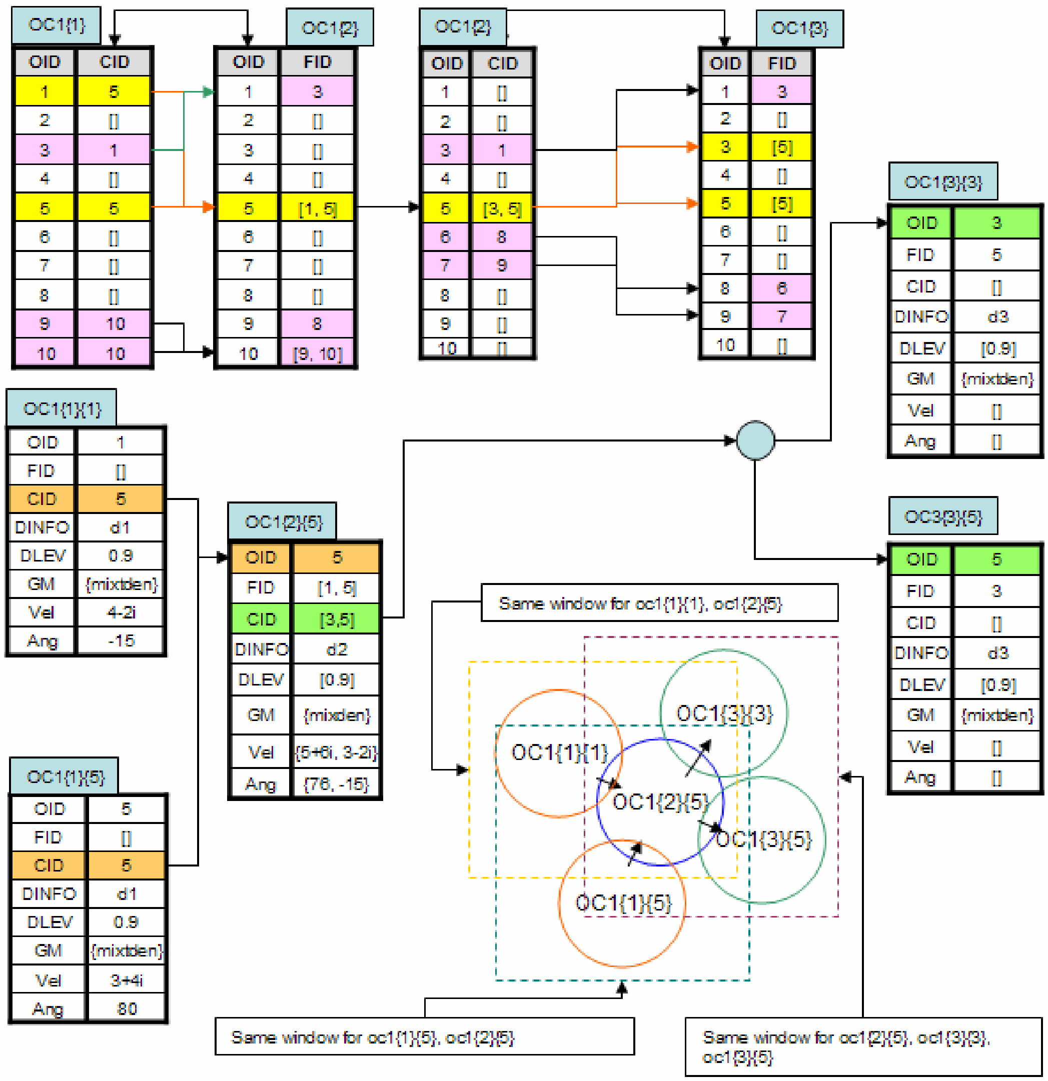


Figure A-27. Schematic diagram of matching storm feature objects.

A.3.7. Storm Feature Tracking

Once all the storm features are identified and matched with storm features for each pair of the NEXRAD images, storm features are tracked, and their dynamic properties are calculated and saved in the text file using the script &task_wrtstix. The &task_wrtstix contains one script file and two functions. The script file &loadcvars loads the necessary workspace variables for use of two functions &mxsavepathcol and &mxwrtstix. The &mxsavepathcol saves the paths of storm features passing over the Brazos County. The &mxsavepathcol contains two main functions: &mxpathfinder and &mxpathoverwsd. It loads the saved record into a record object from the data directory. It calls the function &mxpathfider to find all the paths in the record object. All the paths, however, may not pass over the Brazos County so that the function &mxpathoverwsd is called to find the paths passing over the Brazos County by taking all the paths as input argument. The output of &mxpathoverwsd is saved as the MATLAB data file pathcol[RID].mat where RID is the index for the record. The &mxwrtstix is called in the script file &task_wrtstix to calculate the basic information and dynamic properties of the storm features, and it saves those results into a text file. The calculation is done by the function &mxtrksstix, and saving the results is done by the built-in MATLAB function &fprintf.

A.3.8. Statistical Summaries of the Storm Features Passing over the Brazos County

The statistical summaries are calculated and saved by the script file &task_stixAnal. It calculates the statistics of storm feature residence time, storm feature area ratio, basin area ratio, storm feature intensity, velocity and direction of storm feature, and average storm feature size. The &task_stixAnal consists of cell features. Cell features in MATLAB operate on contiguous lines of code that can be evaluated as a whole in an M-file script, called cells. A cell consists of the line starting with the two percentage signs (%%) and the lines that follow, up to the start of the next cell, which is identified by %% at the start of a line. A cell can be executed by pressing Ctrl + F9 key while a mouse cursor is located in the cell. Each statistical property of storm feature is calculated on the cell features. The figures of statistical results are created by M-file script &task_figures. The &task_figures also consists of cell features. First, the necessary workspace variables should be loaded by executing the cell feature that has the cell title, LOAD WORKSPACE VARIABLES, following the %%. Figure 4-1 and 4-2 can be obtained by the cell feature that has the cell title, FILTERING AND SEGMENTATION. Figure 4-3 can be obtained by the cell feature that has the cell title, STORM FEATURE FITTING. Figure 4-4 and 4-

5 can be obtained by the cell feature that has the cell title, GMM OF A SINGLE STORM FEATURE AND FURTHER SEGMENTATION.

VITA

Name: Janghwoan Choi

Address: Dewberry, Inc., 8401 Arlington Boulevard Fairfax, VA 22031

Email Address: jchoi@dewberry.com

Education: B.S., Civil Engineering, Inha University, Incheon, South Korea, 1998

M.S., Civil Engineering, Inha University, Incheon, South Korea, 2000

Ph.D., Civil Engineering, Texas A&M University, 2006